

Key words: VZV/DNA sequence/genes/homology

The Complete DNA Sequence of Varicella-Zoster Virus

By ANDREW J. DAVISON*† AND JAMES E. SCOTT

*MRC Virology Unit, Institute of Virology, University of Glasgow, Church Street,
Glasgow G11 5JR, U.K.*

(Accepted 23 April 1986)

SUMMARY

The entire DNA sequence of varicella-zoster virus (VZV) was determined using the M13–dideoxynucleotide technology. The genome is variable in size, but the sequence which was obtained comprises 124884 bp. Analysis of the sequence indicated that the genome contains 70 genes distributed about equally between the two DNA strands. The genes are organized compactly, but regions of overlap between protein-coding regions are not extensive. Many of the genes are arranged in 3′-coterminal families, and at least one is spliced. The discerned organization of VZV genes and that deduced for herpes simplex virus type 1 (HSV-1) from published transcript mapping data indicate that these two members of the *Alphaherpesvirinae* are very similar in gene layout. Comparisons of the predicted amino acid sequences of VZV proteins with those available for HSV-1 proteins generally suggest evolution from an ancestral genome, and allow the functions of several VZV genes to be deduced, although limited regions where the genomes differ in functional organization were also identified.

INTRODUCTION

Most people contract chickenpox as children, and many of these suffer later in life the painful symptoms of shingles. Both diseases are caused by the same herpesvirus, varicella-zoster virus (VZV): chickenpox as a result of generalized primary infection, and shingles as a consequence of reactivation of virus which has remained latent throughout the lifetime of the individual. There are no generally available measures for the effective prevention or cure of either disease. Despite such a motivation for studying this medically important virus, our knowledge of the molecular biology of VZV is rudimentary in comparison with that of the other four herpesviruses which infect humans: herpes simplex virus type 1 (HSV-1) and type 2 (HSV-2), human cytomegalovirus and Epstein–Barr virus (EBV). One major reason is that the problems encountered in obtaining sufficient quantities of viable cell-free virus *in vitro* have made the genetic analysis of VZV particularly difficult. Consequently, the most fruitful areas of VZV research have been those which avoid this obstacle. One such area has been the direct analysis of the genetic material of VZV.

The VZV DNA molecule was first shown to have a G + C content of 46% by Ludwig *et al.* (1972). In the years that followed, several VZV isolates were analysed using restriction endonucleases (Oakes *et al.*, 1977; Richards *et al.*, 1979; Zweerink *et al.*, 1981; Straus *et al.*, 1981, 1983; Martin *et al.*, 1982). Significant contributions were made by Dumas *et al.* (1980, 1981), who correctly determined the molecular weight of VZV DNA to be 80×10^6 , and reported the genome structure of VZV and the first three restriction endonuclease maps. These results were confirmed and extended by further structural studies and the derivation of additional maps and construction of libraries of cloned DNA fragments (Straus *et al.*, 1981, 1982; Ecker & Hyman, 1982; Gilden *et al.*, 1982; Davison & Scott, 1983; Mishra *et al.*, 1984). Additional structural features emerged from initial DNA sequencing studies (Davison, 1983,

† Present address: Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, U.S.A.

1984; Davison & Scott, 1985). In summary, the VZV genome is a linear double-stranded DNA molecule consisting of two covalently joined segments, L and S. L comprises an unique sequence (U_L ; approx. 100000 bp) flanked by a small inverted repeat (TR_L and IR_L ; 88.5 bp). S contains an unique sequence (U_S ; 5232 bp) flanked by a large inverted repeat (TR_S and IR_S ; 7319.5 bp). The genome is not terminally redundant, and possesses an unpaired C residue at the 3' end of L and an unpaired G residue at the 3' end of S. Virion DNA contains two major and two minor genome arrangements differing in the relative orientations of the L and S segments; whereas one orientation of the S segment is present in 50% of virion DNA molecules and the other in the remaining 50%, one orientation of the L segment is present in approximately 95% of molecules and the other in only 5%. It has been reported that a small proportion of virions contains superhelical circular DNA molecules (Straus *et al.*, 1981; Kinchington *et al.*, 1985).

As a culmination of our own studies, the VZV DNA sequence is presented in this paper. Thus, VZV is the first member of the *Alphaherpesvirinae* whose genome has been completely characterized at this level. The usefulness of comparisons between VZV and HSV-1, also a member of the *Alphaherpesvirinae*, became apparent when Davison & Wilkie (1983) observed from DNA hybridization experiments that these viruses possess several conserved genes arranged colinearly in the genomes. The hypothesis resulting from this finding, that the two genomes have similar gene arrangements, was confirmed by comparison of the VZV gene layout deduced from the DNA sequence with that of HSV-1 proposed from currently available transcript mapping and DNA sequence data. Consequently, the functions of several VZV genes can be identified on the basis of our knowledge of the molecular genetics of HSV-1, which far exceeds that of any other herpesvirus.

METHODS

Recombinant plasmids. Plasmids comprising *KpnI* or *SstI* fragments of VZV DNA inserted into the *PstI* site of vector pAT153 have been described previously (Davison & Scott, 1983). Additional plasmids consisting of *HindIII* or *EcoRI* fragments of VZV DNA inserted into the *HindIII* site of direct selection vector pTR262 (Roberts *et al.*, 1980) or the *EcoRI* site of vector pUC9 (Vieira & Messing, 1982), respectively, were characterized on the basis of published *HindIII* and *EcoRI* maps (Straus *et al.*, 1982; Ecker & Hyman, 1982; Mishra *et al.*, 1984). For S1 nuclease analysis of the mRNA encoding deoxythymidine kinase, VZV *PstI* *o* was subcloned from the plasmid containing *HindIII* *b* plus *l* into the *PstI* site of vector pUC8 (Vieira & Messing, 1982). All plasmids were propagated in *Escherichia coli* K12 strain DH1 (Hanahan, 1983).

DNA sequencing. DNA sequences were obtained using the M13-dideoxynucleotide technology (Sanger *et al.*, 1980). Plasmid DNA was sonicated and then precipitated using polyethylene glycol to give random fragments 400 to 1500 bp in size. The sheared ends of the fragments were repaired using T4 DNA polymerase in the presence of the four deoxyribonucleoside triphosphates. The fragments were then ligated into the replicative form of vector M13mp8 (Messing & Vieira, 1982) which had been linearized using *SmaI* and treated with bacterial alkaline phosphatase. Ligated DNA was transfected into *E. coli* K12 strain JM101 (Messing, 1979), and clones bearing inserts were identified as clear plaques in a bacterial lawn using isopropyl β -D-thiogalactopyranoside, an inducer of the *lac* operon, and 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside, a histochemical substrate for β -galactosidase. Recombinant bacteriophage DNA was prepared under conditions of good microbiological practice, and clones for sequencing were identified by hybridization of the appropriate nick-translated VZV restriction fragment. Clones were sequenced using pentadecamer primer (New England Biolabs), [α - 32 P]dATP (PB.10204; Amersham) and the Klenow fragment of DNA polymerase I. The latter was obtained from Boehringer Mannheim in earlier stages of the work, but for the majority of sequencing the Klenow fragment was purified from the genetically engineered strain of *E. coli* described by Joyce & Grindley (1983). Sequencing products were separated in 0.35 mm 6% polyacrylamide-urea gels containing a buffer gradient (Biggin *et al.*, 1983). Each gel was bonded to one glass plate prior to electrophoresis (Garoff & Ansorge, 1981), and then dried prior to autoradiography.

Plasmids containing the following VZV DNA fragments were sequenced in their entirety: *KpnI* *t*, *KpnI* *c*, *HindIII* *a*, *HindIII* *e*, *HindIII* *d*, *HindIII* *b* plus *l*, *KpnI* *b*, *KpnI* *i*, *KpnI* *l*, *SstI* *g* and *SstI* *f*. Junctions between fragments were established by sequencing specific restriction fragments from additional plasmids.

Data handling and analysis. DNA sequences from individual M13 clones were read using a Summagraphics digitizer pad with a program written by P. Taylor, and were compiled and analysed using the programs of Staden (1982) modified by P. Taylor for a DEC PDP-11/44 computer operating under the RSX11M system. Open reading frames were identified using the program of Blumenthal *et al.* (1982), and translated into amino acid sequences using a program devised by Taylor (1986). Codon usage was examined using the program of Staden & McLachlan

(1982). Sequence homologies were analysed using the matrix comparison of Pustell & Kafatos (1982) and the optimal alignment program of Taylor (1984). Hydrophobicity profiles were prepared using the parameters described by Kyte & Doolittle (1982). A search of the Protein Identification Resource (release 5) compiled by the National Biomedical Research Foundation (NBRF) for homologues of VZV proteins was carried out using the WORDSEARCH program of Devereux *et al.* (1984) in a DEC VAX11/750 computer operated by Edinburgh Regional Computing Centre.

S1 nuclease analysis of VZV deoxypyrimidine kinase mRNA. Cytoplasmic RNA was prepared by the method of Kumar & Lindberg (1972) from uninfected human foetal lung cells or from VZV-infected cells showing 50% c.p.e. Structural analysis of the deoxypyrimidine kinase mRNA was performed using the S1 nuclease digestion procedure of Berk & Sharp (1978) modified by using 5' or 3' end-labelled probes. DNA/RNA hybridization and S1 nuclease digestion were carried out as described by Rixon & Clements (1982), except that 15 µg RNA was hybridized at 45 °C to less than 1 µg end-labelled DNA fragment isolated from the *Pst*I *o* plasmid. The digestion products were separated on DNA sequencing gels and detected by autoradiography.

RESULTS AND DISCUSSION

VZV genome size

The entire VZV DNA sequence is shown in Fig. 1. It was derived from approximately 1.2×10^6 nucleotides of data, and about 97% of the sequence was determined on both strands. The genome size, from the plasmids analysed, is 124884 bp, and the G + C content is 46.02%, impressively close to the value of 46% derived from buoyant density centrifugation by Ludwig *et al.* (1972). The sizes and G + C contents of components of the genome are as follows: U_L , 104836 bp, 44.33% G + C; TR_L and IR_L , each 88.5 bp, 68.36% G + C; U_S , 5232 bp, 42.78% G + C; TR_S and IR_S , each 7319.5 bp, 59.04% G + C. The significantly higher G + C content of the inverted repeats is a general feature of herpesvirus genomes, and has been discussed previously for VZV (Davison & Scott, 1985).

Five regions of the genome contain tandem direct reiterations of short G + C-rich sequences. One, in TR_S , is a duplicate of that in IR_S , and so the four unique reiterations are denoted R1 to R4 in Fig. 1 and 2. Regions of the genome which vary in size between different virus isolates have been mapped by Straus *et al.* (1983), and correspond approximately to the locations of R2, R3 and R4. R4 (109762 to 109907 and 119990 to 120135 in Fig. 1) has the structure AAAAAAX, where A is a 27 bp element and X is a partial copy of 11 bp of A. Casey *et al.* (1985) reported that the copy number of the 27 bp element varies between virus isolates.

The region containing R3 (41453 to 41519 in Fig. 1) is the most variable in size between virus isolates (Straus *et al.*, 1983). Moreover, fragments containing R3 are particularly difficult to clone in *E. coli*, and those cloned fragments which are obtained may differ significantly in size from the virion DNA fragment (Straus *et al.*, 1982). Thus, the R3-containing clone which was sequenced (*Hind*III *e*) is smaller by about 1000 bp than the estimated size of *Hind*III *e* cleaved from virion DNA. In this clone, R3 has the structure AAAAABAX, where A and B are unrelated 9 bp elements and X is a partial copy of 4 bp of A. Preliminary analysis of an independent clone of *Hind*III *e* which is about 500 bp larger showed that the additional sequence is contained within R3, and that the reiteration contains a complex arrangement of 9 bp elements, including one not present in the smaller clone (data not shown). The sequencing results, and the discrepancy in size between virion and cloned DNA fragments, imply that R3 may contain in excess of 100 copies of the 9 bp elements in virion DNA. Presumably, this highly repetitive structure is unstable in *E. coli*, so that stable clones are obtained rarely and lack many of the 9 bp elements.

In the Dumas strain, R2 (20692 to 21017 in Fig. 1) has the structure ABABAAAX, where A and B are 42 bp elements differing in a single base pair, and X is a partial copy of 32 bp of B. Again, variation in the copy number of the 42 bp elements results in size heterogeneity in this region of the genome in different isolates (P. Kinchington & J. Hay, personal communication). R1 (13937 to 14242 in Fig. 1) is a rather complex reiteration containing four elements: A, B, C and D. A and C are 18 bp in size and differ in a single base pair, B is 15 bp in size and unrelated to A or C, and D is 15 bp in size and consists of the first 6 (or 7) bp of A or C linked to the last 9 (or 8) bp of B. R1 also contains a partial copy (X) of 3 bp of A, B or D. The sequence of R1 was

	AGGCCAGCCCTCTCGGGCCCCCTGAGAGAGAAAAAAGCGACCCACCCTCCCGCGCTTTCGGGGGACCATCGGGGGGATGGGATTTTTCGGGAAACCCCCCCCCG	120
	AGCCTTTAAACAAACCCGCGCTTTTGCTCCACCCCTCGTTTACGTCTGGATGGGACCGTGCACTACCTCCCGGACCTGGGACCCCGCGGTCACCTGCTGCCCCAGCAT	240
	GGATGACGTTGGACCCCATCCCTACCTACCCACATACGCGAGGCGGTGGAGACGCGCCCCCTTACAGAGCGCGAGAGTCTGGTGTCTCCCGGCTCTTTTCTCACGT	360
	GGAGAATGGCACCACCAACGCTTACGATTGCTAGACTGCGCTTATGATGGAATCCAGAGCTTCAGTGGCTTTTCTAAGAAATCGCAATGCTGTACCGGCTTTTAAATCT	480
1	TTTTGGTATTCTACCCCTTACTGCTGCTGGTGGCATTGTTGCGGCTTTTCCCGAGGAACCTCCCACTCACTACATGAACACTGCTCCGGAAGGGGAGGATTTATCTGCTT	600
	* - E R K	105
	GCAGCTTGTGCGCGTGTATGCACAAAGCTATATATGTCACAAAGCCAACGTGCGCATCTGGAGTACTACCCAGTACGTTGCATAACCTGCCATTTCAGTTGCGC	720
	C S T A R T H V V F A I Y T V L A L T A M Q L V V G L V N C L R D M Q M K L Q A	66
	GGAGCGCTTTCTCGGGATCGTGGCTTGGGACATCAACAGTGGAAAGAAGCCGCGTGGTGTGTTGAACGACGAGTGGGACGCGGTTGCTGCAAGCTCTGTATGCTGATA	840
	S A K R R S R P R P V D V L P I L V A P P R T Q V V L P S A N N Q M L E T H Q Y	26
	CATAACACAGAGTCTGTATGCTATCAGATCCCGAACACCTTCCGGTACCCCATCTCCGATACCTGGACATTGGCGATCCCAAAATATAATTAACAGGATTGCTTATACATTT	960
	M F Y S D T D S D S E R V G E P V G Y E S V R S M	1
	GCTACAGCTTATATAAATTTATGTCGATACATCTTAAGTGCATCCGATCGTTATTTATACATTGCTGTCACGTGAAAAGACTGTTTACCAATAAAGTTCTACAAAAATGCTTTA	1080
2	TTGGGTGTTTGTAAATAGCTATTATCGTAACCCACCCCGTAAATCATAAATGCTGTAATTTCTGAGACACTTGCATATGGCATGTTCCCGCATTTATTATGGCTCCACTCTGG	1200
	M H V I S E T L A Y G H V P A F I M G S T L	22
	V R P S L N A T A E E N P A S E T R C L L R V L A G R T V D L P G G G T L H I T	62
	TGCGTCCAGTTTAAACGCCACCCCGAGGAAATCCCGCTCAGAAACGCGATGTTTATTACGAGTGTTCGGGGAGAACGTAGACCTGCCAGGCGGAGAACGTTACACATTACCT	1320
	C T K T Y V I I G K Y S K P G E R L S L A R L I G R A M T P G G A R T F I I L A	102
	GTACAAACCTATGTAATTTATGGCAATATAGCAAAACCCGCGACGCTTAGCCTTCCCGCTAATAGGGCGTCAATGACGCTGGAGGTGCAAGGACATTATTTATTTGGCGA	1440
	M K E K R S T T L G Y E C G T G L H L L A P S M G T F L R T H G L S N R D L C L	142
	TGAAGGAAAGCGATCCACACGCTTGGGTATGAATGTGGTACGGGCTTGCATTACTGGCTCACTATGGGTACATTTCCTCGCACACAGGTTAAGTAACAGAGATCTCTGTTAT	1560
	M R G N I Y D M H M Q R L M F W E N I A O N T T E T P C I T S T L T C N L T E D	182
	GGCGGGTAATTTATGATATGCATATGCAACGCTTATGTTTGGGAGAAATCGCGCAAAATACCAGTGAACACCTTGTATAACGTCGACGTTAACAATGCACTTGACAGAGACT	1680
	S G E A A L T T S D R P T L P T L T A Q G R P T V S N I R G I L K G S P R Q Q P	222
	CTGGTGAAGCCGCACTTACCAGCTCAGACCGACCCACTTCCCAACCTAACAGCCCAAGGAAGACCAACAGTTTCCAACATTGCTGGAAATATTGAAGGATCCCCCGCTCAACAGCCG	1800
	V C H R V R F A E P T E G V L M -	238
3	TCGTGACCGGTTAGATTGCGCAACCTACGAGGGGCTATTGATGTAATCACTAAATAAATACACCTTTTTCGATTGTACGATTTTATTTAAATGTGATGTCATGCGCCG	1920
	- L G G	177
	ACAGCCGCTCGGGCTTTTCCCCACATACAATGATGCTATGCTCGGATGCAACGCTCCCGGAGAGGGGATTTACAATGACAGTATACCAATAGCCGCCAGATGT	2040
	V A A R A K G G C V V H D Y A E S A G T W C E A S F P S K C H C H Y G I A A L H	137
	ACACCCAGCTGTCGGGATCCAGCATCTGCTGAGTTGCGGCGTGAAGGGTGCATGCAATAGGGTGTATAATTAGCCATTTCGGTAACAGCTGTTGGGAATTTAGGAGGCTGCA	2160
	V G L Q G S E L M W Q Q T A A S F P A D C L T N Y N A M E P L L R Q S N L L S C	97
	AACGGCTGTAGGTCAACATACATTGGGGATTCAGATGGTGTATCTCGACGTCCAAGTCCAATCAAAAGCGTGAATATCATGCGCCGCGCATGTTGCTCGAAGAGCACAATACCTC	2280
	F P Q L D V Y M P S E S P K D R R G L G I L F A H L D D A R G C T A R L A C L R	57
	TTAACACCGTACAGAGGGATGGGCTGCTGATGTGAGTTGGCAGGCGATGTCACGTTGTTTCAACGCCAGTGGCGGTATAACTGTGTAAACGACGCCAAGGGTCAGGTTTAA	2400
	K Y G Y L P S P T P A H S N A P C T M T T E L A L P P I V Q T F S A L P D P K L	17
	TTCACCTGGATGGGTTGACTGCTTTCGGAAGCTCCCGTTGTATCCATTAATTAACAGTTCGGTACACGCTGGTGTGTTTACCAGAAATCAGAGCGGAATGCAAGATATTGGTTT	2520
	N V R I P Q S S E S A G T T D M	1
	GAAAGCAATGTAATCCCGCCATATATCCCAACGCTGCGCTTAAACCTCCCAATATTACATTTTATTAGTCTTTTATTAATAGAAATCACATAAACAATGATAAAATCAAGGG	2640
	TGGTGTATAATGATTAATAATATAATGATATGTTTACAAGCATGAATAGGATTTACTATTCTAACAGGTAATATGCTTAATGATTAATAAATACAAATAGTATGTTTGACAAG	2760
4	CATGAAAAGGATTTTATTTTATGAGTAAAGGTACTACCTTAAATATTTACGATGAGAGGGGCTCAGAAAGATGCCCCCAAGTTGAGAGGGATCATTAACACGACAC	2880
	- C N F T S C K F Y K G Y P R A D S L H G A W T S L T C E V R G C	421
	ACTCGCGTGGTGGGTATTAGGGCTCTAAACACCGGCGAGACATGACCGGGTGTATATTCTGTAAACACTTGAACGTTACAACCTGATATCATATTTCCACAAATTTAGAGCCAC	3000
	E R Q H T I L A E L V G A L C S G P T Y E Q L V Q V N C S I D O Y E V F K S G R	381
	GGACAATATATTAGCAATGCGGGCAATCAACAACATATAAGTAGTAATACAGTGATACCTAAACGTTGCTGGCGCAACAGTTCGGGAGAGTACGAGACCCCAATCGTTGT	3120
	V V I N A I R A I M V F M Y T T I C T I D S F R Q Q R L L E P L T R S G L D N D	341
	CCCTGTTTGAAGAAGACATCTTACAAAGGCCCGAGCTTAACTTTAAATCTCCAAAGTGACTTCGAGTTGCAACAATGGGATTATTTGTGTAGATGGCAAGTTTTCGCGCTA	3240
	R N L L L C R V F P G L K L K L N E L L S K S T A V I P N N T Y I P L N K A A L	301

VZV DNA sequence

1763

ACATTTTAAATCCAGTTAATCCAGCTCCAGACTCCAAAGCTTCAATCAAGATTCTCCACGTATGACTCTCTACGCAACGCGGGCAATACGTAGTCCATTTATATGACTCAA	3360
M K I M T L L E D A S E L A E I L S E G R I V R E R L A R A I R S D N K Y S E F	261
AGGTACGATAAAGTTATGTCCTGACATCAATCCGCGCAAGATGTGTTTATCCCGGAAACATCCACGGAGCCCAATGATCACCCTCTGTATTGTGGCATATCGGA	3480
T R Y L E H G V L M L E P W S T N Q K I G P F C G G S A W S D G E Q I T A Y R V	221
CTACCAAGTTTTCATTTTCATTAATGGCGTACCGAGTCAATGGTCACGTGGCTCCCGGGTGGAGACGACTTCAATAGCAGGCGCGTAATTCGATCGACGGGATATCATACT	3600
V L K E I T E D L H R V S D I T V S A G A T S V V E I A R G T I R O V P I D Y E	181
CTTTTCGAATACGCTCTCGGCGGGCTCTCTCTGGAATCGCAACCTGTACGATTCGTATGTGATCATTCTTCTCCCGTGGTCATTGAGGAGGGCTGTAGGACGCCGTC	3720
K R I R E R R A D R K S F R L R Y S E D H T Q D N R E G T Y M A P P T T P R R R	141
TTGATTTGACAGGGATCGATCAGCGTGTCTTGAACCTTGAAGTGTATAGATCGGATGATCGTATGTCGATTCCTTCTCCCGTGGTCATTGAGGAGGGCTGTAGGACGCCGTC	3840
R N S L S R D R H K R S S Q T N Y S R S S R R H G R E I R A Y G T E V D R G G H	101
GATGGTTGAATCGGGTAAATCAACAACCAAGTTTTCGGGCGATTTGGTGGTAGCTTTCACGCTTCGGTGGCTGTTGGAATACCGTGGATTATATGCTGTATCGAGTACGCT	3960
H N S D P L V V L T K P R N H H Y S E R R G H R R K S Y R P N Y A T D A T R E	61
CCACATACAGTCTAGACGTTGTGGAGTCTCGCTGGAGTGGAGCAATAGCTTCATCATTTGCCAATCGGTGACTTCCAATGCAAGTCATCGAAGGTTGCTGTGGTACCAAT	4080
V Y V T R S T S D E G P T S G I A E D N A W D T V E L A F D D S P E D P L L N	21
TCATAAGTCTTCAAAATAGTAGACGCTGGGTGGTGGAAATGAAGCAGAGGCCATGGCTGCAAAATATCTGACAATTGCGTGTTCAGTGGCTGTATCTTCCGCAATGTTG	4200
M F D E C I T S V D P D T P I S A S A M	1
TAGAATTTATAGGCTACCCCAACCCGCAATGGGCGTGTAGTCACATGATTAACTCTTGGGAGTTTTCACCTTCCCAACCAAGCTTACCTGCACCTTTGTCGTAATGCATAAA	4320
- H K Q S N E S E G F L S V Q Y R Q E Y H M F	319
ATAACCACTGCTATAGCAATATGACGATATGCAATATTAAGCAAGGCGGACATTACTGTAGCGCAACATGTTGTGCATATACCAGTATTCGCCCGGATTTGATGATTAA	4440
I V V A I A F I V I Y F C K I A L G S M V T A C C T T C I G R I G G T N I H N L	279
ATGATTATCCTTGGTGGTGTGTTGCTAACATAGATATAAGCTCTACTATAGCGAGCGTGCATACAACAACCGGCGAGATCCGAATGATGTGGGGTATAATAACGCGCATGGTGT	4560
H N D K T P K P R V Y S I L E V I A L T C V V W A L I R I Y T P Y L L A C P T	239
ATATGCAACGCAAGCTTAAAGACCAATACATCCAGATGATATGAGCGATAACCTCCAAAGCATCAATAACGTAACACCTTTATGCATATATAAAAACTTATAGGGTCAGCATT	4680
Y A V G L T L L V I C G S S I H A I V E L L M L L T V G K H M Y L F S I P D A N	199
AAATACCTTACTCATACATCCCGTGCATGGAACATCACAACAACCTTGCCAACTTTGTATATGGGTAAACGAAGAATGTTGCAAAATACCGGTGTTACGTAATTCAGTGAATA	4800
F V K S M G D R R M S V D C L L R A L K T T P Y G L L I N S I V R T V Y N L S Y	159
TGATGTGGGGGATTAATCAACAGATGATCGGAATGGCCAAACATACGACGATTCGCTGAAATTTGTAATACATACCATATACAACCATGCAAAAAATCATTTTATGCTGCAC	4920
S T P S I L E C S S R F P G F M R R I R R F Q L Y N G Y V F W A F F I M K L Q V	119
GCACCAAAAAAAGCGTGAATATCGTGTCCCAAGCAATTCGAATTTGTCTGCAAGGTTGAGAAATAGCGGTTTTTACCATAGTATCTCTGATAATAGATTTCCCGGCGAGCT	5040
C W F Y A H C N R T G L V I R I K D H L P T S I A T K V M T D G S L L N E R C S	79
GTATCGTATCCAGATAGGCAATCAAAAAAGTGTGAGTGTACAAACGTTACATATATAAGAGAGTGTGTAAGACCCCATACAACCGGTCACCATTAATCACCCTGGTGCATA	5160
Y D Y G S L G D L F T S H N V F T V Y I L S N N Y S G W V V P G G N I V T T A Y	39
CACACACTGATGTTCAACTTACAGAGCGGTATACCATAGGGTAAAAACAGCATGTCGCGTAAGTAGACACATAATTATAAATGTTCTGTCTGATTCCTAAAGCCTGATGACCCG	5280
V C E H E F K V R A T Y M L T F V A H G S L L C M I I F H E T K I G L A Q M	1
TGGAAGATGCAATCAAGCAGATGATGATACACGCTGGTGTAACTCGAAGTTAAATTTGGAATAATTAGGTACTTCTAGAGTAAAGATTGTATGCATGCGATTGCTATCGCACTT	5400
- S S T L N P Y N P V E L T F I T H M R N S D C K	1060
TGTAGCAAAACATGTTGTGCAAGCAATACACAACGGTTGTGATGATCCACTCGCAGAGACAAATGTCGGGAGCCGTTCTTCTCCGCGATGGGGATATCGAAGACAAGTGA	5520
T A F C Q Q A L S I C L R N H D V R L S V F T R P A T R G G R N P Y R L C T F	1020
CCCTTTTGTCCGATATGAGTGAATTAACACCGAGTCCCTTTTGTGCGGATACACTTGTATGATGTTAAGGTATATTCGCGATCACGCCCGGGGAAATGAACAGCAATATGCTCCAC	5640
G K T G C I L Q F L V N D R K I A I C K S S T L T Y E R D R G P F H V A I H E V	980
AATAGATTCTAATATTGTGCTGCGACAAAGGCTCCAGTGAATGCGTCCAGACAAGTACCCCGCGCTCTTTAGAGCCTTTGTTAAAGATATTGCGGGAGCTAAATATTGTTT	5760
I S E L I T S D V F A E L T F A D L C T V G R E K L A K T L S I Q P S S F I Q K	940
ATTACGCGCAACCTTACGTTCAAAAACCTGCGGTATGTTCCCAAGGTTATGTAATAAATTTGCACTGGAACATTCGACTGCGGCTTGAATGAAATGAAAGTTTCCCGGTTTCT	5880
N R A V K R E F F E A Y E G G L N H L I F Q V P V N S Q P R S H F H F N A P N R	900
ATGTGATGTCACAAACGTAATATATCAATACACTGCTCAGGTACACATAAATGGAGTAGTTGTCCAAACCGCGTCCCTGTGGTGTACTTGGAGAAAAAGGAGCTTAAACT	6000
H S T V F A L I D I C Q E P V V Y F P L L Q G V A T G T T T V K S F F P L R L S	860
ATGTCGCTGCTATAAACACCAATATCTATAAACGAAAAAGTCCCGTAATACGGAACATATATTCACAAATCCCGTCCAGCAACCCGTTGCTGTAATATTGTGCAAAACCCCTT	6120
H G H S Y V G T D I F S F D R L Y P G I Y E V F E R E L L V A Q Q L I Q A F G K	820
TAAAGTGAAGACCCCACTACGATAGGATTTGGGATTGGTACGATACCTGAAACCTATTTTCTTTACAGTTACAGGGTAGAGTTTTCATGCAAGTTTTCATTGTTGATACATC	6240
L T S S G V L A Y P N P I P V C V R F G I K E K C N C P L T E H L M E N N S V D	780

GGCGTGTATGGACTTCAGACGTTGCTGTGTATCAAAAACCATACATCCTCTGTATAATCTCTCTACACAGTGTATAATGCCATTTCTATGTAAAAATCGATGTCAGAAATG A H T H V E S T T Q T D F F N V D E T Y N E E V C T Y L E G N E I Y F D I D S H	6360 740
GCTGGTTATATCAATAATTAATCATCAACACCTCAACGGTAGGTCAGGACATGAGTTTATAAAAAAATACATGGGCTTTGTTAGGGTTTACCACGGCCTTTGGAAAAAGTAA S T I D L L N D D D L V E V T P E P C A T K Y F Y C P D K N P N V Y A K P F L L	6480 700
TTGCATGGCGTTAAAAATACCATGACGAAATGCTCGCATGCCGCGATGTAATAACCAATGGGATGGGTTTCTTATATGAAAGTCTACATCAAGTATGAGGTTTGTGATTATAAGATT Q M A T L I G H R F A R M G A H L I G L P I P K R I H F D V D L I L N T I I L N	6600 660
TGTATTAATAGCTCATTCTGTTTATATAAGCTGATCTTGGGTATGTTGTGATAATTTAGAAACGTTTAAACAGCATAGATAATAGTAAGTCAACTGCATATCTCGTAGTGA T N F L E N R N I Y L Q D K P I N S S I K S V N K V S T S L L L T L Q M D R L S	6720 620
AGCGGCAACAAAATTAATGATGTTAATTTGTTAAGGTCCTCCGCAATTAATCGAGCCTCGTGGGTAAGGTAAACGGTTTGTATTGATGACCAGTATCATTAGCAATAACAGCAAA A A V F N C P N I Q K L D E A I L R A E H P L T Y R N T I S S N T D N A I V A F	6840 580
TGCTTGGGCGCGTGAGGCAAGGCTACCGATATACAGGCAATGCTCCAGTTACCTCAGAAATGGCGATGAGGCGCTTAATGGAGTTTATAACTCAGGATGATACATCATGTGTGGC A Q A G H P L A V R Y V P M P G T V E S H G I L A E L P T K Y S L I S V D H T A	6960 540
TATCCAGTGGCAGCAGAGAAAACAGTAATAGTTTGTAAATCCCGGGCTCGTATCAAAACAGTACGACCACTTGGTTAGGTATCGTTTGCAGGTTGGCTGCTGTAACGCCCTC I G T A A S F F L L L K T I G P S T D F G T R G S Q N P T D N A F N A A R L A E	7080 500
CGCGGAAACACCCGAATCTCAAAATAGACAATTCGTAACACCGGGTGATTTGAGGGAATAGTGAGGAGACCATCATATGGAGTAAATGTTTTCAATGTTTTCCACAGCAGAGT A S V G S D E F N S L E D F G P P N S P I T S S W G Y P S F Q K E I N E V R R T	7200 460
TAGCGTTGTAGCTAGGTCACATACGCTATAAATCTGCTAGGTTTGGCGCATACGTAAGACTTAAGATATGTTTGTAGTAATGTATATTTATGTCGAATCTCAGGTCAGGTTTCACT L T T A L D C Y G I F K S P K A A Y T L S L T Y T K T I T Y K H G I E P G L E T	7320 420
GACATCACAATTAACGTTCTTTTATATAGTCACGATGTTGAGACGAGAACGTACATGATTAATAAATAGCAGTAGCTCTTTTCCAGGTTGGATGTTTAAAGAGGACGGGTTT V D C I V N K K I Y D R M N L R S R V H N F F N A T A R K G L N S S K L L V P K	7440 380
ATTCACAAAATCTGAGTATGTAACCGCTTGTAGGTGGCTGCGATCTGTTCCGATGAAACATTCAAAATGTCCAGATAAATAAATCAACAAATTCAGGCTCGGAACTTAAAGCC N V F D S Y T V A Q L H D A I Q K R H F C E F H A L Y I Y D V F E R D P V K L G	7560 340
TTTCTATGCTGGTAATATACTCCGATACGCTGATTTCCGTTGTGCTGTATGTTGCTGTAATAATGTACGATAGCAGATTTTGGCTGTCAGAACCTCGTATATGTTGAGGA K R D N T I Y E S V A H I E T T D T H I R Q L I Y S L A N K A T L G R T Y T S S	7680 300
ACAACAAAATCGAAAGTTTATCAAAAGACAACAGTCCGAAATATTGTACCCACTACAATAGGTAATGCCGGGACTTGGTAAGTTAAAAACAAATCTTTAATGCTGTAAAGTCATA C C F M S L K D F S L L D S I N Y G S C N P L A P V Q Y T L F L D K I A Q L D Y	7800 260
TAAGGGGTTTCCAACTGATTGTAACCTGTCGTTTGTAAACAGTAATAGCGTGTAGGCAACACTAGCGTTTTCAGAGGGTCCAAATCGAAATATACCAAAACGGGAGCATCC L P T E L T N Y S T D T Q L L Y R T A L V L T K E S P G F R V I Y W F P S C G	7920 220
ATACCCCGTAGAGTGTGATATGACGCAATCTGACGTCGTAATGGGATATAATGATGTTAGCTCCTGACGACCAACGAGTTTAACTAATCTGACAGGTTTGCCTCTGT Y G W Y L R R Y A A L V Q R E Y H A Y L S T L E Q R G V S K K V L K C L T A E T	8040 180
GATGATAGGCGGTTGTCGATAATCCCTTCGGTTTAAATGGTGTGTTGTTACCATCAGAGTTTGTATAACTTCGAGGAATGTCAACGCTCCGATATACATAGGATATCAGATAT I C L G N D S L G K R N L H N T T V M L T Q I V E S H I D F T E S I C L T D S I	8160 140
TATATGCGGATTTAGGGGTGCTCATACCAATCAACGCTTATATAAAGCTTAAATCAAGTTTGGGTTTAAACAAACAAAAATATAGGCGAGACCGGATCTACATCTCAGTTGA I H P N L P A G Y W L A K Y L A K F D T Q T K F C C F I P W V R S R V D G T S	8280 100
AAATCCACCAATTAATAAAAAATAACGTTGACGTCCTACTACAAAATAATGATTATTTGGTTTCTTCATCGTTTTCAGTTACTTCACGTGGGCTTGTAGTTGGGATTACTTGCCT F G G I L Y F I V N V D R S C F L H M I Q N E E D N E T V E R P R K T P I V Q T	8400 60
GATCTCTCCCTCCATTTTGACAAGAGCTATCAAGTCGGAGTCCAAGTATAACTCACCACATACAGAGGTTCTGTGCTTATCTGCGGTAAGCAACAGCAGGAGGAGAT I E E R G N K V F V D D L D P T N T Y S V V Y L P E T S I Q G T L L L L S H S I	8520 20
7 TGCACATCCCTTTGCGCAATAATAACCGAATCGTGGTTTGGAGGATTATCCATAGTTCAATCGTTGGAAAGCCAGTCAATCATGACAGGTTGTGCCAGCTTATGTGGATATG A C G K T A F L L R I T P K S S K D M	8640 1
A R I P T E E P S Y E E V R V N T H P Q G A A L L R L Q E A L T A V N G L L P A C T C G A T A C C A A C T G A A G A G C C A T C T T A T G A A G A G G T G C G T G T A A C A C G C A C C C C A A G G A G C G C C T G C C G C C T C C A A G A G G C T T A A C C G C T G A A T G G A T T A T T G C C T G C A C	8760
P L T L E D V V A S A D N T R R L V R A Q A L A R T Y A A C S R N I E C L K Q H C T C T A A C G T T A G A A G A C G T A G T C G C T T C G C A G A T A A T A C C G T C G T T G G T C C G C C C A G G C T T T G G C G A A C T T A C G C T C A T G T T C G T A A C A T T G A A T G T T T A A A A C A G C A C C	8880
H F T E D N P G L N A V V R S H M E N S K R L A D M C L A A I T H L Y L S V G A A T T T A C T G A A G A T A A C C C G G C T T T A A C C G C G T G G T C G T C A C A C T G G A A A C T C A A A C G G C T T G C T G A T A T G T T T A G C T G C A A T T A C C C A T T G T A T T A T C G G T G G C G G G	9000
V D Y T T D D I V D Q T L R M Y A E S E V V M S D V V L L E K T L G V V A K P Q T G G A T T A C T C G G A T A T A T T G C G A T C A A C C T G A G A A T G A C C G T G A A G T G A A G T G G C A T G T C T G A T G T T T T T G G A G A A A C T C T T G G G G T C G T T C T A A A C C T C A G G	9120
A S F D V S H N H E L S I A K G E N V G L K T S P I K S E A T Q L S E I K P P L C A T C G T T G A T G T T C C C A A C C A T G A A T T A T C T A T A G C T A A A G G G A A A T G T G G G T T A A A A C A C C A C C T A T T A A T C G G A G G C A C A C A A T T A T T G A A A T T A A C C C C A C T T A	9240

VZV DNA sequence

1765

IEVSDNNTSNLTKKTYPTETLQPVLTTPKQTQDVQRTTPAI 251
TAGAAGTATCGGATAATAACACATCTAACCTAACAAAAAACGATTCGACAGAACTCTCAGCCCGTGTGACCCCAAAACAGACGCAAGATGTACAACGCACACCCCGCGATCA 9360

KKSHVMLV - 259
AGAAATCCCATGTTATGCTGTATATAATATTGAAATAAAAACTAAAAACGTTCTG6GTATGTTTTATTTGTATATAAAATTAACACATTGCTG6GTG6GTG6GTATTACATTTAA 9480

YGTITTAGTAAAAATCGACATCGTTTGTCTTTATCAGTTGAACCAATCCACGCGTTCGCCGTTCGCTG6GTG6GTATTAAGATCTAACGTTTTAGTAAAAATACCATTGTACACCC 9600
8 H K L L F D V D N T E K D T S G F G R T G R E S P T A I L D L T K T F Y M Q V G 357

GGTATGCCACATTTACCGCGGATAGCATAGGAAATGCAATATTACTTAAACGTTGTGTTTAAAGTGTATTTGGGTGTTG6GTATTAACAGGACCTGTGCAAGACGATCTCCCGT 9720
P I G C K G R I A Y P F A I N S L V N H K L H I Q T N H D I L L V Q A L R D G T 317

TTTATACGTATGTCATCCCGTGAGATTATATACGTAGAATTTACAGTGTCTCTGACGGCCATGCGGTGGACACACGATAATGCTGATCGGCTTTTCGATGATCTTCAAAAAATA 9840
K I R I D D G T L N Y V Y F K C H E G A P W A T P C V I I G S R S K S S R G F I 277

TAAGCGTTTATCTCGGATGTGTAGTCCGAGCTCTTATAATCGGTAAGCAATTTTATAAATTCATCTTTTAAATATAGGTTATATG6TACACAAATATCATATCCCGCTCT 9960
Y A N I S P H Q L D W D R I I P L V I K I F E N R K L Y L N Y P V C I D Y G A D 237

TCTTGGCGTTTGGATGATATGTTGTAGGTTAAGGAAACATCGATATG6TATTCGACGAATCCCTATGTAAGGTTGCGCTGCTGATCCGTTGAAATATCAGCAATTCAGGT 10080
E Q R K P N I I H K Y T L P V D I H Y E A S D R H L P Q G Q Q V T S I D A F E P 197

ATAACGGGTTTTCATATTTGACGCGAGTTGATAAGGTTGAACCTGTATCGATTTAAAAATG6ATCCAGATGTTAAGAACGTTTTTGGGAGAGGCGACTTTGTCTTAATTT 10200
I V P K E Y N S P S N S L P Q V Q I S K F I P D L H K L V M K P L L R S Q R L K 157

ACCGGGAACAGTATGTTAAATGTCGCGGTAATAACAGGTTACTCTGCGCGGTAATAACAAAGGCGTGAATTTACTCTGTGTAACCCGATCAATAACCTCGTTGCGGACAAAA 10320
V P F L Y I T L H G P L I V T V G P R Y L L A S I V G R Y G A D I V G N A V F 117

AAATGTCTTCATCAGCAAGGCGAGTCTTTGATTGAATTAACACAGTGGTATTCATGGGAGGCGCGACTTAACCAACAGCTCCAATGCTGATATAAAACCGCCCGTGT 10440
F N D E D A L A T D K C Q I L L L A Y E N P P A S K Y L L E L Q Q M Y F G G R T 77

ACAGATTTTCAGATGGCAGTTCGAGTTTCTGTTG6TTCGAGTAACAACGGTTGATGTCGACTTACTTTATCGTCTAACACGCTTGCAGCGTATCTGCACATTAGGTTGAACCTCT 10560
V S K E S P L E L K K H N R L L L P Q H R S V K D D L V C Q L T D A C E P Q V E 37

ATTAATTTGATCTTTTAAACACCGATTCGGAATAGTTG6GTACAAAACATACCTGTATTTACTGCGGTTTCCAGATGGGATCAATACCCTTCGTTTCATATTAATAACGATG 10680
I L I T D K L C R N P I T Q S C F M D G T N V A T E L I P D I V A E N M 1

CAAAATTTATTTTTTGTGAAGACAGTGGGAGCAAACTTTCGAGAACGGAATTTTGGCATGCCAGTGTTCG6GTG6GTG6GTATATATCGAGGATCAATGATCACCACCTT 10800

TTCTCTACGCATCCCTTTTGGGGGTGTGTGTAGCCCTTATTCGTTAGCTTATCATGCGGTGTTCCG6TATTACTCGTCTGTATTACGACGACGTTGTAACCCGTTTGCCTATA 10920

9 M A S S D G D R L C R 11
AAAGGGGAGGCGGTGATAAGAGGGCCCTGTTAATACGCGGTCTGCCGTGTTGGATATTCACGACCCATCGTTTATTTACGTAATGGCATCTCCGACGTTGACAGACTTTGTCG 11040

S N A V R R K T T P S Y S G Q Y R T A R R S V V V G P P D D S D D S L G Y I T T 51
CTCTAATGAGTGGTGTGTAATAACAGCCCTAGTTATTCGCGCAATATCGAACCGCGCGGGAAGTGTG6GTG6GTAGGACCCCGATGATTCAGACGACTCGTTGGGTTACATTACCAC 11160

V G A D S P S P V Y A D L Y F E H K N T T P R V H Q P N D S S G S E D D F E D I 91
AGTTGGGCGGATCTCTCTCCAGTGTACGCGGATCTTTATTTGAACATAAAAAACGACCCCTCGCGTACATCAACCAACGACTCCAGCGGATCGGAGATGACTTTGAAGACAT 11280

D E V V A A F R E A R L R H E L V E D A V Y E N P L S V E K P S R S F T K N A A 131
CGATGAAGTAGTGGCGCCCTTCGCGGAGCCGTTGAGACATGAACGTTGAGATGCTGTATGTAACACCCGTAAGTGTAGAAAACCATCTAGATCTTTTACTAAAAATGCGGC 11400

V K P K L E D S P K R A P P G A G A I A S G R P I S F S T A P K T A T S S W C G 171
GGTTAAACCTAAATAGAGGATCACCAGGAGCGTCCCCGGGAGCAGGCGAATGCGAGCGGAGACCAATTCCTTCAGCACTGCACCAAAAACCGCAACAGCTCGTGGTGGG 11520

P T P S Y N K R V F C E A V R R V A A M Q A Q K A A E A A M N S N P P R N A E 211
TCCTACGCCATCATATAACAAACGCGCTTTTGTGAAGCGGTCGCGCGTACGCCCATCAGGACACAAAAGGCTGCCGAAGCGGCTTGAATAGTAAATCCCCCAAGGAATACGCCGA 11640

L D R L L T G A V I R I T V H E G L N L I Q A A N E A D L G E G A S V S K R G H 251
ATTAGACCGTTTGTAAACCGGAGCGTTATTCGATTACG6GTGATGAGGTTTAAATTAATACAGCGCTAATGAAGCAGACTAGGTTGAAGGAGCATCGGTATCCAACGTTGACA 11760

N R K T G D L Q G G M G N E P M Y A Q V R K P K S R T D T Q T T G R I T N R S R 291
TAATCGAAAACTGGAGATTACAGGGGGGATGGGTAATGAACCTATGTACGCACAAGTTGTAAGCCAAAAGTGAACGGATACACAAACGACTGGGCGTATAACTAATCGAAGTATG 11880

A R S A S R T D T R K - 302
GGCCCGTCTGATCAAGAACTGATACGCAAAATAGGATATAATTACGAGTAACGGTTTACCCG6TATTATGTAATAAATAAACGTATAAAAGACAGTGTG6GTTG6GTTATT 12000

ATAAATGTGATTATATGTCACATATTATAAATGTTTAAATAGTACCAGTGGTATTATGAACAGTTTATATCAGTTGCTACCAAAACACCCCATTAGACGCGG6GTTTGTATAAG 12120

10 M E C N L G T E H P S T D T W N R S K Y E Q A V V D A 27
GGAATCGCTTATTAACTAAGATTTTACTCTAATAGTATGGAGTGAATTTAGGAACCGCAACATCTTAGTACAGATACGTTGAATCGTATGTAACCGGAACAGCGGTTGTGGACGA 12240

F D E S L F G D V A S D I G F E T S L Y S H A V K T A P S P P W V A S P K I L Y 67
TTTGATGAATCGTGTGTTGGTGTAGCATCGGATTTGGATTGAAACGTCGTTATATTACATGCAAGTAAACGTCGCTCCGCTCGG6TGTAGCCCTAAAAATTTATAT 12360

[illegible]

VZV DNA sequence

1767

F Y P T E R E M L A S F F T L Y V T L G G G M L N W I C R A T A M Y L A A P V H 590
T T A C C C C A A C G G A A A T G T T A G C G C A T T T T T A C G T T G T A T G T T A C T C T T G G T G G A G A A T G T T G A A C T G G A T C T G T A G A G C A A C T G C A A T G T A T T T A G C T G C T C C T T A C C A T T 15360

S R S A Y I A V C E S L P Y Y V I P V N S D L L C D L E V L L L G E V O L P T V 630
C C C G T T C G G C T T A C A T C G C G G T G T G A A T C T G C C C T A T T A C T A T A T C C C G G T T A A T A G T G A C C T G T T A T G T G A T T T A G A G G T A T T A C T G T T A G G C G A G G T C G A C C T C C C A A C T G T T T 15480

C E S Y A T I A H E L T G Y E A V R T A A T N F M I E F A D C Y K E S E T D L M 670
G T G A T C C T A C G A A C T A T T G C A C A G A A T A A C C G A T A T G A G G C T G T T C G C A C A G C A C A A A T T T A T G A T A G A G T T T C C C G A T T G T T A A G G A A A G T G A G A C C G A T T A A T G G 15600

V S A Y L G A V L L L Q R V L G H A N L L L L L S G A A L Y G G C S I Y I P R 710
T A A G C G C T A C T G G G G C G T T T A T T G T T A C A A C G G G T T T G G G T C A T G C A A A T C T C T T T T G T T G C T T C T C C G G T C T G C G T T G T A C G G A G G A T T C A A T T A C A T C C C C G A G 15720

G I L D A Y N T L M L A S P L Y A H O T L T S F W K D R D D A M Q T L G I R P 750
G T A T T T A T A G A T A T A A T A C T T T A A T G T T G C A A G A A G T C C T C T T A C G C T C A C C A A A C T T T A A C A T C C T T T T G A A A G A C C G C G A T G A T G C A A T G C A A C T T T G G G A T T C G A C C G A 15840

T T D V L P K E Q D R I V Q A S P I E M N F R F V G L E T I Y P R E Q P I P S V 790
C A A C G A C G T T T A C C C A A G A G A A G A G A G A T A G T C A G G C A T C A C T A T A G A G A T A A C T C C G T T T T G T G G A T T G G A G A C C A T C T A C C C G A G A A C A G C C A T C C C C T C C G T G G 15960

D L A E N L M Q Y R N E I L G L D W K S V A M H L L R K Y - 819
A C C T A G C C A A A A T C T T A T G C A A T A C G A A T G A A A T C T G G G T T G G A T T G G A A A A G C G T A G C A T G C A T T A C T A C G A A A T A T T A A G G G T T G T G A T T T T T C A T T A G G A T G A A A A G 16080

A A C G T T T C C T A G C C A C C C A C A A A G G A G T T T G A A A A T A A A T C T G T T T A G A C C T T A A A A T T G T T G T G T G T T G T G T G G G G G C C G T G A G A G T G A C C T T T A C A A G A T A A T T 16200

M F S R F A R S F S S D D R T R K S Y D G S Y Q S F N A G E R D L P T P 36
T G T C C A T A T G C A A T G T T T C T G G T T T C C G C G T C C T T T T C A G C G A T G A T A G A A C C G T A A A T C T T A T G A T G T A G T T A C C A A G T T T A A T G C C G C G A A C G T G A T T G C C C A C C 16320

T R D W C S I S O R I T S E R V R D G C L I P T P G E A L E T A V K A L S E K T 76
T A C C C G G A C T G G T G T C A T T T C C C A A G C A T A C C A G C G A G C G T G A G G A T G G A T G T T A T T C C A A C G C C G C G A G G C T T T G G A G A C G G G T A A G G C T T A T C T G A A A G A C 16440

D S L T S P V L Q S T E R H S V L L G L H H N H V P E S L V Y S C M S N D V H D 116
C G A A G C T A A C A T C C C G G T T T A C A A G T A C G A A A G A C A C A G T T C T G T T G G A T T A C A C C A A A T A A T G T T C C T G A A T C G T T G G T G G T C T G T G T A T G C T A C A G A T T C A T G A 16560

G F N Q R Y M E T I Q R C L D D L K L S G D G L W V Y E N T Y M Q Y L K Y T T 156
C G G G T T A T G C A G C G T A T A T G A A A C A A T C A A A G A T T T G G A T G A C C T G A A A C T T T C T G G G A T G G A C T T T G T G G G T T A T G A A A A C A T A T T G G C A G T A T C C A A T A C A C C A C 16680

G A E V P V T S E K V N K K S K S T V L L F S S V A N K P I S R H P F K S K V 196
A G G A G C C G A G G T A C C G G T A C T C A G A G A G G A A A T A A A A G T C A A A T C C A C G G T T T G T G T T T C A T C C G A G T T G C C A A A A C C A A T A C C A G A C A T C C T T T A A A T C T A A A G T 16800

I N S D Y R G I C Q E L R E A L G A V Q K Y M Y F M R P D D P T N P S P D T R I 236
T A T A A A T C G G A T T A C C G G A A T A T G C A G A G C T A C G T A G G C G T T A G A G C T G T G C A A A G A T A T A T G T A T T T A T G C G T C C A G A T G A T C C T A C A A C C C A G C C G G A T A C A A A T 16920

R V Q E I A A Y T A T G Y G M W L W F L D V D A R V C R H L K L Q F R R I R G 276
A C G T G T A C A A A A T T G C G G C T A C A C G C C A T G C T A C G G T G G A T G T T A T G T T T G G A C G T T G T G A C C C A G G G A T G T G C C A T C T C A A A C T C A A T T T C G A C G G A T T C G A G G 17040

P R A S V I P D D L L R R H L K T G P A V S A G T G V A F I L A A T T A S A L T 316
G C C G C G C G C T G T T A T T C C A G A T G A T T G C T A G A G A C A T T A A A A C G G G C C T G C G G G T C T A G C G G C A C A G A G T T G C G T T A T T T A G C A C A A C A C T G C C A G C G C T T A C 17160

A L L R I S V L W R K E E W R D G L N G T A A A I V A A V E L I T L L H H H F Q 356
T G C G C T T T G C G A T T A G T T A T T A T G C G A A A G A A G T G C G G A T G T T A A T G G A A C C G A C G T G C A A T T G T T G C G G G T T G A C T T A T T A C G C T T T G C A C C A C A T T T T C A 17280

Y L I N M M L I G Y A C W G D G G L N D P Y I L K A L R A Q G R F L Y F A G Q L 396
A T A C T T A A T A T A T G A T G C T T A T T G A T A T G A T G T T G G G G A T G G G G A T T A A C G A T C C T A T A T A T A T A A A G C G C T A C G T G C C A G G A C G G T T T A T A T T T G C G G G C A G T T 17400

V R T M S T H S W V V L E T S T H M W F S R A V A Q S I L A H G G K P T K Y V A 436
G G T C A G A C A A T G C A C A C A C A T T G G G T G T T A G A G A C C A C C C A T A T G T G G T T T C C C G G G C G T G G C G A G A T A T T T A G C A C A T G G G G T A A C C C A C A A G A T A T A T G C 17520

Q V L A A S K R Y T P L H L R R I S E P S S V S D Q P Y I R F N R L G S P I G T 476
T C A G G T T C T T G C C C A G T A A A C G G A T A C T C G G T A C A T T A A G A C G T A T A T C C G A A C C A T C G A G T G T G T C T G A T C A G C G T A T A T C G T T T A A T C G A C T G G G A T C C A A T A G G A C 17640

G I G N L E C V C L T G N Y L S D D Y N A S S H V I N T E A P L N S I A P D T N 516
A G G T A T A G G A A T T G G A A T G T G T T T A C C G G A A A T A T T A T C T G A C G A C G A A A T G C A A G T T G C A T G T A A T T A A T A C A G A A G C A C G T T A A C A G A T A G C A C C G A T A C A A 17760

R O R T S R V L V R P D T G L D V T V R K N H C L D I G H T D G S P V D P T Y P 556
T A G A C A G C G A C T T C C G G T T T A G T C G T C G A C A C G G G T T G A T G A C T G C C G A A A A A C C A C T G T T G G A C A T A G G C C A T A C G G A G G A T G C C A G T G A C C A C C A C G A T C C 17880

D H Y T R I K A E Y E G P V R D E S T M F D O R S D L R H I E T Q A S L N D H 596
T G A T C A T A C A C C G G A T A A A G C G G A A T A T G A A G G T C C G G T C G G G A T A A T C A A C A C A A T G T T G A C C A A A G A T C G G A T T A C G T C A C A T A G A A C C A A G C A T C T T A A T G A T C A 18000

V Y E N I P P K E V G F N S S S D L D V D S L N G Y T S G D M H T D D D L S P D 636
C G T A T A G A A A T A T A C C A C C A A G A A G T G G G T T T A A C T A C T T C A G A C C T G G A T G T G G A T A G C C T A A C G G T A C A C C C C G G A G A C A T G C A T A C A G A C A G T A T A C C A G A 18120

F I P N D V P V R C K T T V T F R K N T P K S H H - 661
T T T T A C C C A C A G C A C T C C G T T A G A T G T A A A A C C A C G G T A C G T T A G G A A A A T A C G C C T A A G A T C A T C A T T A A G T A C A G C G G T A A T A G A T A G T A T G G A C T A G G C A C T T T G G C 18240

	GGTCATTCCCAACCAGGTAAATATGGGGATTGGGAGAAAATAGCTATTGCGTATTTCTGTTCAATAATGGAGCTGGTTATTTAAAGGTCTGATTGGTGGGTATATAA	18360
13	AGGAATTACTCCTTTAAATTTTACTTAATGACCACAATATCAAGTGGTGGTTGATTTAAACGATTATTACCGGTACCATGGGAGACTTGTGATGTTGGACAAAGGTGCCGGGTTTAA	18480
	Y L T G E L Q V L K Q V D D I L R Y G V R K R D R T G I G T L S L F G M Q A R Y	53
	CGTAAACGGGGAATTCAGTACTTAAACAAGTGGATGATATTTAAGGTATGGAGTTCGGAAGACGATCGAAGCAGGAATCGGAACGTTATCTTTATTTGGAATGCAAGCTCGATACA	18600
	N L R N E F P L L T T K R V F W R A V V E E L L W F I R G S T D S K E L A A K D	93
	ATTGCGGAATGAATTTCTCTTTAACTACAAAGCGTGTGTTTTGGAGGGCCGTCGTGGAGAGGTGTTATGGTTTATCCGGGTCACCGGATTCAGGAGACATTTGGAGCGGAATA	18720
	I H I N D I Y G S S K F L N R N G F H K R H T G D L G P I Y G F Q W R H F G A E	133
	TACACATATGGGATATATACGGATCGAGCAAAATTCCTAAATAGGAATGGCTTCCATAAAGACACACGGGGACCTTGGCCCAATTTACGGCTTCAGTGGAGACATTTGGAGCGGAAT	18840
	Y K D C Q S N Y L Q Q G I D Q L Q T V I D T I K T N P E S R R M I S S W N P K	173
	ATAAAGACTGTCAATCAACATTTTACAGCAAGGAATCGATCGCTGCAAACTGTTATAGATACAAATTAACAAACCCAGAAAGCCGACGATGATATATGCTGTGGATCCCAAGG	18960
	D I P L M V L P P C H T L C Q F Y V A N G E L S C Q V Y Q R S G D M G L G V P F	213
	ATATCCCTTAATGGTACTACCTCCATGTCACAGTATGTCAGTTTACGTTGCAACCGGAGATATCTGCAAGATACAGAGATCGGGGATATGGCGTTGGGGTACCGTTCA	19080
	N I A G Y A L L Y I V A H V T G L K T G D L I H T M G D A H I Y L N H I D A L	253
	ACATTGCTGGATATGCACTCTTACCTACATAGTAGGCGATGTTACAGGACTTAAACCGGAGATTTAATTCATCAATGGGGATGACATATTTACTTGAATCATATAGATGCTTTAA	19200
	K V Q L A R S P K P F P C L K I I R N V T D I N D F K W D D F Q L D G Y N P H P	293
	AAGTCAGCTAGCTGATCCCAAAACCTTTTCCCTTGAATTTATTCGAAATGTAACAGATATAAAGGACTTTAAATGGGAGATTTTACGCTGTATGATATAATCCACACCC	19320
	P L K M E M A L -	301
	CCCTAAAAATGGAAATGGCTCTTAATGGATTTTAAATGTTGTCAGACAGTAGATGTTGCGAATGTAATAAATGATATACAGAGCGGTTTGGTGGTTCTGTTTATGAACAG	19440
14	- S C	559
	CAACGGATGCATAGGTTGGGATAACGCGATAAGACCAATGTCCCAAGGATAGATATCACACCAATTATAACTGCTACAACGGAAATGATGGGCTAGGATAGATCATGATGTA	19560
	C R I C L T A I V A I L G L T G L I S I V G I I V A V V S F T T A Y T S A D Y T	519
	TAAACGGCGAAAACGGAGGAATTTTTAGGGTAACCATCTAGATGACACGAATAGGTGATAGGTCCTCGAGTTCGATGTTGGACAAGAACTTTCATGTTTACAAACCGTTTGT	19680
	Y V A S F P P F K K P Y G D L H C S Y T I P G D L E S T P C S S Q M N V F R K N	479
	TGATCACACCCAGTAATCTCATGTTTCTGTTTAAATGGGAGATGTTAAACCCACCATACGAATGTACAACGCCACGTGGCAACATTTTCCGCTACATACTATGTGCTCATCA	19800
	Q D C V G T I E S N E H N I P S D N V W W V F H V V G R P V C K A T C V I H G D	439
	ATAATACCTATAGACACGTTGGGAAATGGATAGACGTCAGGGGTAAACGACGAGAATATTTATATTAGAGCGCCATCCGGAATCCATAAATACATTCGATGGCTGGGGTGGG	19920
	I I G I S V N P F P Y V D P T V V A S Y K M N S V G D R I W L V N C Q I A P P P	399
	TAATCCATTGTTTCTGCTGGGATTCGTACCGCCGAACATACTAAATATCCATTGGCATATTTCTGTTGATGCTGTTATAAATTTTTCCGATGTACCAACCTTGAAGTC	20040
	Y D M Q K Q Q P I R V A S V S F L G N A Y E Q I A D T I F N K G I N G F R S T	359
	CACCGAACACGTACCGAGTGGGATCTAATCTTGTATACGTTACAGTAGGCTGCTATGCTGCTGGTAAAGCATGGATCGCCGACACCGGATATTTGGAGATAATACGTTGTA	20160
	W R V R V S H P P Y Y K S V N C Y A A Y T Q G T L V P D G V V T I N P R Y T T	319
	ACTGTAATCTGTTGCTGATATGAGCTTCTAGTTTTTGTATTAACGACTCGCAATATACGTTCCCTCGGTGAGCATCCATAGATAAATTTGTACAGAAAAACAGAGTTGTT	20280
	V T I S H E S I V N K T K T N V V R M I Y T G E T T A D M S L I T V S F O S T T	279
	TTAATCATCTGGTATTACATAATTTCTTACGCTGTGTAATATCTCAGGGTTGTTTATTAAGTTTAAATCGGCACTGTTGCTATATAACATAACCGGTAATCTGGATCGCTATTAAC	20400
	K V D P - I V Y N E K A H T F I E P N N I L N L D A S N S Y L M V P L D P M R I L	239
	GCATTGCCAGTTGACGGTGGGATCTAATGGTGAACGCTAAACCAACTCAATATGAAGATCGGGGCGTATAAGCGACTTCCACCTTGTATATTTGAACCTTCCGGATCAAGAA	20520
	A N G L Q R H P D I L H R T F N V E I H L D P R I L S K W R T I N S G E P D L S	199
	TATGTTATATGTTTGTGTTGCTGCTTAAAGGCCCTGTTGTCGGTCTGTTAGAGCATGTAACAAGGATGATAAATGTTGAAAAATAGGATGATGATTCCCGCTGAACG	20640
	Y Q E Y T K Q Q Q K F A A Q Q G T T L R M Y C P M I F T H F Y P I S Q I G G H V	159
	CATTGTATATTTTCATATAGAAAAGGTGGTGTGAATGTTGGGTGTTGGCTGCGGGATCGGGCTTCGGGAAGCGCCGAGGTGGGCGGACGCGGGATCGGGCTTCGGGTAGCGGC	20760
	C Q I N E Y L F P P Q S H Q T N A A P D P K R S A A S T P A V A P D P K R T A A	119
	GAGGTGGGCGGACGGCGGATCGGGCTTCGGGAAGCGCCGAGGTGGGCGGACGCGGATCGGGCTTCGGGTAGCGCCGAGGTGGGCGGACGCGGGATCGGGCTTCGGGAA	20880
	S T P A V A P D P K R S A A S T P A V A P D P K R T A A S T P A V A P D P K R S	79
	reiteration R2	
	A A A S T P A V A P D P K R S A A S T P A V A P D P K R S A A S T P A V A P D P K	21000
	I	39
	CGGGTAGCGGCGAGGTATATAATTCAGTTATCTACGGGTGGGTTGAGATTGAGTCGATAATTGTATACAGCGATCGTTAAATTAATTTATTTGTATCCGCTTCATCTGTT	21120
	R T A A S T Y L E T I S V P T P Q S E T S L Q I C A I T L I L N I Q I R K M	1

VZV DNA sequence

1769

TTTATTGACATCCACGCTCCCCTTAAATAAAGATTAAACACCCACCGGGAATTTAATGATGGAACGTTTTTTTCGACATTGGGAATAATAAAACGGCTTTTGAACCTTTAA 21240

15 AACCTTTATTTATCTCGATTACGATACATATGTACCATAGATAGATTTATTATAATATAACACACGCGATATCTTTAGTGATGAGATGCCATAAAACAGTCAATAGGT 21360
- S V Y T G C L Y C L N I I I Y V C V H Y V K T I H S A M F C D I P 374

TTAACGCTTAGTCTCATCTGAATACAGCTCAACCCCGCCGAAGTGTGATGTTAGAATTATAATAGTCCCCATGAAATGCCGCAATGTTACAGCTATACCCGTACCGAGGTC 21480
K V S L R M M Q I C T L G A A V T S T L I I I A G W S I G A F T V A I G T V S T 334

GTGTATATAATAAATACCATAGGTTTTTTTTTTTGTATATAAACCGGCAAAACCTGTAACCCAAATGCTATAATAGCTCTCTATTGAACTGCTAACGTTACTTGTGAAGT 21600
T T Y L V I V M L N K K E Q Y L V A F G Q L G F A I I H G G I S V A L T V Q T L 294

TTGATAAATGATTTAATTTAATATATGTGAGATTGCCACATTAATGGGTAACATATATAACACCGGGGTATAACAGACATTATACGAATCTTTAAACACGGCTTTAAGGGTC 21720
K I F H N L K I I H S I A M M L P T V I Y L V P P I V S M I R I G K F V R K L T 254

CGGGAAGTCTTCGATGTCACATCTCTCCCGGTCATTTTGTATATAACCGCAAAACCTAAATCTGTATAAGTGTAAATGCTTATGCGGATTTTACGATATATACAGTA 21840
R S S E R H D C V R G R D N Q T Y V V A F G L D T Y T N L Q K H R N K R Y I C T 214

TCTGCAATCGGTGGCGCATCGACAATGAACTAGTGTGACATAGATATACACAATCCAATAAGAACCTCATATTTACTGACATACATATAAAATAACGGTTAGTAACCTCCC 21960
D Q L D T A A D V I S V L T V I S I C L G I L V E Y K S V Y M Y L I V T L L G G 174

AACCCAGTTCACATCATACATAAAATAAATATGCGTCCATTTGAATGCGTAACAAAGTGTAGTAATGATGACAGCAGCCACTGTTCCGGTAATCGCGATGGAAT 22080
L G T G L M M V Y F Y I H P G N F T T V F N Y Y H I H V A A V T G T I A S I S I 134

CCCAGTAATCTACAATGGAAGATCCCGGATATGGGCAACCAACCGCCATAACACAGCAAAACCCACACGACCCGCTGCAAACTGCTCCCAATTTGCTAATGTGCTAGA 22200
G L L E V F P L D R S I P C G V A M L V A F G L V V V T Q L C R G L K A L T R L 94

AATTCACGGATGTGGCAATAACCCGGAACGACATCAACCCATAATGTTGACGCGAGCTTCGAGAGCTGATTTGTAATTAACCCGACGTAACGCTTCTGT 22320
F K V S T P M L G S V V I L G M I T A N V A A E C V H Y Q L I L G S T I V S A Q 54

AGTECCAGGAGAAACACCGCATGCTGAGGTTATTGACACGAATACATCTTGTAGGGTTCCGACACATCTTGGATTGATGAGCGCGGATTAATCTCTGCTAACACCCAGG 22440
L G Y L F L R S A S T I A C S W C E Q P N G V C G Q I S Q A R I L E R D L V G L 14

TTTTCATCATGACAGCTCTTACCATTACCGGCATGCTTAAGTTAATAATCAAAACAAATAAAATGTGTTCTATGATGACACAGAAGTTGTATGTAATATAAGCAAAA 22560
N E D H V A R E G N V A M 1

16 GTTGACATTTTAACTGTACATATTACGTCAGATTACGTGATAATTCAGAAATAACGAGGTTCTCGAGGGTCCACTGGAGGAGCCACACAATATTCGGAATTCGATCCCTCT 22680
- K V T C I V D S E R S L E S Y D L T G A P D V P P A V C Y E R I G I G E Q 372

GCCATGTGGTTTCGGGAGTTTCCCCCATTTTATTCGGTATTTTTCGTTCTTTTGTAAATTAATGCGTCTTTTTTAAATGAGTTCATCTTCACAGATTCATGTCG 22800
W T T E P L K G G M K N G T N K E N R K T L L N R R K K L P P E O K Y S E M N A 332

CAATAATGTCATGAGGTTAATTTTCTTAAGGCTTTGGGACTTAAGAAGCTTGCAATAAAAGAAATGACAGGGTGGGAACTGGATATACAATCAACATGGGGAGTTAG 22920
F L Q M S T L K E K L D K P S L F T A Y F F S H V P A S R Q I Y L G V M P S N T 292

TTAAGGCGAGATAAAATAATATAACAGCTCTATCCGTTAACTTAAGATTTGTACGCGAGAGGAATCCACTGTGTGTTTCAATAATACTCCAAATTCACGATATCCCGC 23040
L A L Y F N I Y C T E D R T L K L N Q V A S R F G S H T E L L V G F E R M S G S 252

TGCCATAACACATTAATGAAGTCTTTTGAATTTGTGATTGAGCTTAATTAATATGAGTGAAGCTTCTTCCGTTATATCCAAGGAAACATTAATGAGATAAACACCCC 23160
G Y V V N M L I R K S N T I S R I L N Y P T P K S G N I D L S V N F S I F G G G 212

CGCGGCTGGATGATATCCGTCGTTGAGATGAAGCATGTTGTAACCAAAAGTTTAAAGTAGTGGCTGTAACGAGGTAATGATCGGTTTAAAGCAATGCTTATATCTG 23280
A T Q I Y M D T A T L I F C T T F G F T K L L R Q L R T F Q D R K L C I S I D P 172

GAGTTAGATTGGAACATCATTTGATAACAGCGAGTTCACGTTTACAACCTGTTTGAACATTGTACTGTATCTGACCAACATACCCGGGCGTTGCCATACCATGTTTGA 23400
T L N P F M M T Y C A L E R K V V Q K Y C Q V Q D D P G C D G P R Q M V M T Q I 132

TAATACCTCGCTCGGGGGTTGTCGGTAAATTAACCGTGTGGGTCGACGATCTTTGTATGCGAAACGCTCAATAAGCGAGGACCGTCCCTCGTGGCGGAGTA 23520
I S R E P P Q G T F K F Y L R T P T S P D K T H R F A D I L S S R G E J A A L V 92

CAACCTTCTCGGCGAGTCAATATATCTGTTGCAACATATTTGCCGGTATAGGAATATACAGTGTGTTTCCAACTACAGTGAATAATATCTTCTGCTGAATATTA 23640
V M R P G T M N Y Q D F M N A P I P I Y L Q E T E L S C H I I L G E D S F I L I 52

TAGAACCTTCTAGTCTTAACCAAGGATATAGACGAAATAAACAGTAAGCGTTTTTTCGTTAAACAGCTCTGGCGATTTCTGGGGGCTCAAAACCGCATGCAATTCATGT 23760
S D R L R N V L P S I S S I L G T L T K E T L V A R A I E P A D F G A H L E M D 12

CCAAAGCATGCTGTACGCGACCTCAATCCATAATTTACTTAAATGTTTACTATAGAAAAGTAATCATATGTAACACACGAGTTTCGTTAATGTTTGTAAACCGATCC 23880
L A D D T R S R L D M 1

GGTGACTAAGTACATAACAGCATGATTTGATAGTACGCGCATGGGAGGGAACATTCCACGTGTCCAATACAGGGGGTGTCTTAATAGGGACTGTGCAATAAATACGTA 24000

AGAAGTTACAGATTTGATGTAATGTTGTCATAAAAAATATGATACATTATACGCTGTGTAATTAACACAAGATCAGTGAAGAATTACTGAAGCCGCTGTGAACCTTTCACAA 24120

17 M G L F G L T R F I H E H K L V K P S I I S T P P G V L T P V 31
GACGATATAAAGTGTGTAAGTGTATTGATGGGCTCTTTGGACTGACAGCTTATCCATGAACATAAAGTGTAAACCCAGCATCTTCAACGCCACCGGAGTTTAAACCCCGT 24240

	A V D V W N V M Y T L L E R L Y P V G K R E N L H G P S V T I H C L G V L L R L	71
	GGCGGTAGCGTATGGACGTCATGTACACATTGTGGACGTTTATACCCGTGGGTAACGCGAGAAATTACGCGACCATCTGTAAACGATACATTGTCTGGAGCTTATTGGCGCT	24360
	L T Q R S Y Y P I F V L E R C T D G P L S R G A K A I M S R A M N H D E R G T S	111
	ATTAAACCAACGGTCATACCTCGATATTGTATGGACGTTGTACAGACGCGCCATTATCACGTGGAGCCAAGGCAATTATGTACGCGGCATGAACACGATGAAGGGAACCTC	24480
	D L T R V L L S S N T S C S I K Y N K T S E T Y D S V F R N S S T S C I P S E E	151
	GGACTTAACCGGTCTTACTATCATCCAACACATCATGTTCTATCAAGTATAACAAACATCGAAACATATGACAGTGTGTTTCAAACTCTCCACGAGTGTATCTTAGCGAAGA	24600
	N K S Q D M F L D G C P R Q T D K T I C L R D Q N V C S L T S T M P S R G H P N	191
	AAACAAATCCAGGATATGTTTTTGGACGTTGTCCACGACAACTGACAAGACGATCTGCTGCGGACCAAAACGATGCACTTACCTCTACAATGCCATCCGAGACATCTAA	24720
	H R L Y H K L C A S L I R W M G Y A Y V E A V D I E A D E A C A N L F H T R T V	231
	CCATCGATTATACAAATGTGTGCAAGTCTTATTAGATGGATGGGATGCAATACGTCGAGCGGTTGACATTGAGCGGACGAGGATGCAAACTTATTTCATACGGTACAGT	24840
	A L V Y T T D T D L L F N G C D I L L D A I P M F A P V V R C R D L L Q Y L G I	271
	GGCTTTAGTATACGAGATACGATTACTCTTCATGGGCTGTGATTTTGTAGATCAATCTCTATGTTGCTCCAGTAGACGATGCGGATTGCTCAATATTAGGAAT	24960
	T Y P E F L V A F V R C Q T D L H T S D N L K S V Q Q V I Q D T G L K V P H Q M	311
	TACATACCTGAATTTTGGTTCCTTTGCTGTCGACGAGTTCATACAGTACCACTTAAATCTGTCAGCAAGTATTACAGGATACCGGCTGAAGTTCACATCAAT	25080
	D T S T R S P T Y D S W R H G E V F K S L T V A T S G K T E N G V S V S K Y A S	351
	GGACATTCACCGGCTCCCCACTTACGACTCGTGAGACATGGCGAGGTTTCAAAAGTCTTACCGTAGCCAGTGGGTAAACAGAAACGGAGTCCGCTTCAAAATATGATC	25200
	N R S E V T Y D A S W A L N L L P P S S S P L D N L E R A F V E H I I A V V T P	391
	TAACCGATCGGAGTGACAGTAGACGCCAGTTGGGCTTAAACCTTCTGACCCCTCATCTCCCATTTGGATAATTTGGAACGCGCATTTGTTGAACATATAATCGCGCTGTAATCC	25320
	L T R G R L K L M K R V N I M Q N T A D P Y M V I N T L Y H N L K G E K M A R Q	431
	ATTGACCGCGGTCGCTTAAAGTTAATGAACGTTGTAATATTATGCAAAATACGGCAGACCCATATATGGTTATTAACACCTTATATCAACTTAAAGGGGAAAAATGGCTCGCCA	25440
	Y A R I F K Q F I P T P L P L N T V L T K Y W N - *	455
	ATACGACGATTTTAAACAGTTTATCTACTCCACTCCCACTAAACACTGTATTAACAAATATTGGAATTAACACACATAGAGCGACTTAATGGTTCATTGTTTATTGTC	25560
	CGTATATACATGTTATAAATCGTTTATCATCTGTCGCCGATAAGATGTACTGTCTCTCAAAAAAATTGTTTATTCGCAATCATAAATGCAAGTGAAGTCCGAGTCCGGAGG	25680
18	- L D N I V T G A Y S T S H R E F F N T N K D A I M F A L P F D S O P P	272
	TGGGTTGTTAAATAGTTTGGTACATTAACGCTGATAAAAGCTGTCGCGCTGAATTCACGATTGTGTGTAATGATGACGTTCCACCAAGGGTTTGGGTGATGGGATTAA	25800
	P T N F L K P V N I A S L L R D A S F K V Y Q T I A D V N V L R T K P A H S K L	232
	AAACGCACCTGCTTTCAACGGTTCGAAACAGTGTGATGTTCTGGTATAGCGGGTTTTCGGGTACATAGTTATGTATATACACAGTGGGTGATGATGGCTTCATC	25920
	F A C E I E V A E S F L Q H I R T I A P K E P V Y N N Y I C C S A S T H I A E D	192
	TCGGCTTAAAGTCTGTTAAATGACAAGTTACAACAAATAGTCGGTATTGCGTAAATGCAATAGCGCGAAGCATGATAAAAAAATGCCCTCTATAAGATCAITAGTATATA	26040
	R S I L D N F Q C T V V F L G N N R L Y A I A A F S S V F F I G E I L I M L I Y	152
	TTTTCTGCAACGGATGGGTTGCTGCTACCTTTTCTCAACCATGTAATTTTGTGGATGACGGAATTATTAAGTACATTTACGATTGTACCGGCAAGTATCCCTCT	26160
	K E A V S P N D R V K E E L W Q V K Q Q I S P N N I T V N V Y Q V R L S E D G R	112
	GAACAACATTAGTTGAATTTGACTATAGACGCGCGTGACAACCTCGATGCACTTGTTCATGTAATGGTGAATATCTCTTTGGGAAAGAGTGGGTAGAGAGCCCAAT	26280
	F L M L Q I Q S Y V R A H V Y E I C E Q E I Y Y H H I D K Q S F L Q T L S G L N	72
	AACATTTACAGATCATCTGCGCGGATAAAAAATGAAAAAATCTGTAGAATATTAGTTCATCTTCCGTTAAACAGTCCAAGTATTGATAATCATCTCAATGATAAATCGCTTC	26400
	V N V L D D A A S L F T F I F R Y F I L E D E T L C D L Y Q Y D D E I I F D S E	32
	TAACCAACGATTGAAATGCTCAGGCGACGTAATTTGTATATCTGGACACTCCGGCTGTAATAAATGACTGCAATCTTCTGATCCATTTTGAATAGTTTCCCGTGTAAATTA	26520
	L W R N S I S L A R L N N I D P C E P R Y F F H S C D K Q D M	1
	TAAAGCACAACGGTACAGGTTAATTCGCTCCCGCAACAGTCCGCTGTCGAGCTTACGAAATTTACAGTAGTACATACCGGTTTAAAGGCGGCTTATAGGACGATATAAGCAA	26640
19	L A C S T C T L E G G A F L G S N T A K R I K C Y Y M G T K L G A K Y A R I L L	736
	ATTCATTATTTGGAGCGGGAATTGTCGCTGCGGCTTCTCAATAAATAGTCATTGATGACTTTGGTCAATAATGGCGGCTTCTGACACATATCAACGAGATCCTCTG	26760
	N M I K S A P I T G D P R E E I F L T M S Q S Q D I F P A R E A C M D V L D E Q	696
	CTCATTTCAACGCTGTTTATATTTAAGAGTGGGTGACTATTAGATAAACAGCCAAACGAGTATTACTGACCATGTTTCTCAAGTATGTTATAACTTCCAGTCTTTTT	26880
	E Y E F A T K Y K L L P H S N S L C G F S R I V S W Q N K E L I N I V E L R K E	656
	TTACATGAATACATATCTTATGTCGTCATAAGGTCTAAGTGGGTCTAAGTAACCTACCCGAGGTGGTACCTTACTAACATATTATTATAAATGGAGAAACCTCCTGCA	27000
	E C S Y M D R L E D M L D L N P R L L E G S T T V K S F M N N Y I P S F G E S C	616
	CTCCGTTACCTGTCAGATGAACTGTGGGCTAACGCTAAGAAGTGGGTGATAACCCATAAGGCAATATCATCTCGAGGATACCATGTTAAATCAATAACTTATGCT	27120
	E T V Q A S S V T P M L A L F Q S N Y L G Y A C I D D R L T C W P L D L Y S I T	576
	AGAAAACCATCTTGGTGAACCATCCCTAGCATATTTACTTTGCGTAAACCTTAAACGGGCTAAGCGGCAATCTTACACATTTCCATGCTTGTTCATTGCTCATACACAT	27240
	S F G D Q H L W G K A Y K S E T F G K F P A L G G I K C M E M S T K M T E Y L M	536

VZV DNA sequence

1771

TAATCCGCTATTTGTACATTTAACGCTAGCTGGTTGGGAAGTTAAATCAATCTAAGCGGAGACAAGTTGTATGTAAACCTTGATGCCAATGCCAAGTATCGGTTGTTTTTAC 27360
LEA IQVNLRRAPQSTLDFGLRLCTTHLGGIGIGLSRNNKV 496

ACCTTTACATGATTTTTACATGGAAAGTCCAGCCGCGCAGGACCCGTTAAAAAATAACAGCTGTTCTGCTGTAATTGAAGGTCGTTAAATTAATGACACTGGCCCTTTGGA 27480
GKCSKKCPFNGAALVGNLFIVTTRATLQLDNLNFSVPK 456

TAAGCACGTTGAAGATTATGCTGGCAAGATTACATACGCCATGTTGATGAGCGTCTGCTTTTGAACAATTCGCTACACAATTTGACCCGCTGATAGCATTTCTTGGGATTTCAT 27600
LCTTLNLSALNCVGHQHADAKQV IETCLNSGTIAN 416

ATGATAATTACGATTACAGGCATCTTGAACATTAAGAGGGGCTTCTGTTACAGCAGCACTGCGTATGATTGTAATGCGATCTTGAATGGGAACAGAAGAACGCTAATCCTTC 27720
HYNNRCADKFM LFP SGTV AASRIITFAIDQIPVSSVGLGE 376

TCCTCTAAACGTAATAGGTTGAAGTGAAGTCCCTCCCGTGAATGTTGCAAGGATTCGCTGTTATCAAAAAGAGTCCACTGAACATTACTAGCCCTTTAGATAGCTTAGGTA 27840
RELRLLYTSTFAEGHLTRLIDARNDFLTMQVNSAGKLYSLY 336

TCCTTCAAAAATAATCGGGTCCATAAACAACAAATATGTTATCACATCGAAATATTCATCAGCAACCAATCCACGCTGCGCAACAGTTGTAGATCGACGTCATG 27960
REFFLDPTW LCCFINDCRFIEDRVLMGR TALVTQLDVHWP 296

TYCTATGTAACACAACTCCAGTGGCTGCTTCAACATCACTGTTAATGCCATAACCATGCAATCAAAAGTTTAAAGTGAAGAAGACCTTTCGTTGATTTCCTGAGGTATTA 28080
EIVVCVGT PRECD SHIANVMCDLLKLVALLGK TQNETPIL 256

ATTCAGACTCTGTAGAGAAATCCCACTCCACCTCGACTTGTATACCGTTCACATCGCTGTGATGCTCGAAGCTCTCCCAACAGTGTGATTCCGGGTCATTAATAACA 28200
NLSQLSIGVGGRSQ LVTGV DGTIARVARGV TI SEPDMLYC 216

ACTGGCGTTGCCCGCTCTCGACCTAAAAACATCAACCGGTGTAGCCGGACAATTTCTGACATGCCAAGCTGTGAAAAATACCCGACAGACATCAGTCCATGTATAACCATC 28320
SATAGTERGLFMWVPTAPVIKQCALATFFVRCVD TMTYGD 176

ATTTATCCGGGAATAGAGTTGCGATTAGGCAAGTTACGATTCTGTTGTACGCTGCGCCAGCTTAAAAAGAAATGGCAAGGACTCAATTTACCTTCTCTAATCTAGT 28440
NIGPILTAIKPLNVIETTYTAALRLFFQCLSELKGEELKT 136

TAAATAAAGCTTCGCTACTTAAAGCAGACTGTAGTCAAGGATAGCTAAGCGGGGATTGATCTTTCAAAACGGTCTAATATAGCCGACGAATTCGCTCCCGCCCTCAAT 28560
LYFDEYKLA S QLGLTALAPYQDKLFPEL IARRIEDRRGEI 96

TGCTGGCGGACTCGGGAGTTAAACAGAGAAATGGGAAGTCAACCACTTTCATGGAACGGATCGTAGGTTAATACGGCAATGGATAAGTCTCCACAACATCGGTACACTCGCTC 28680
AQRVRPTLCLIPSTLWTEMSVSR LNI RCHILEGCCRYVRE 56

ATCTGTGCGCTACCCGCTTAAGTTGTAGAGATAGTCTAATATACCTTAATTCACCGGTGTTGATTGCGGGGGAATGATGATTCCTGTAGCCATGTTGACATAATCG 28800
DQRTVAKLKSVITSIYEMLEVP T TSEPP I IYEKVGHQCLR 16

GTTTATAATGTCATGAACGTAATTAATAATCTTTTGAATCCATAACGGATAACGATTATAGCTCCGGAATAAACCTTTAAACCTTAACCTACAGCTGAGTAGTCTACAATATG 28920
N I I D H V T N F I R K F E M 1

TAGACTCCCTATATATGTTAGTACGACGCTGCCCTCCCAAGTATATAATACGCAAAACCCAGCTATGTTAAATTCAGTTATTTTACATACATGCTTTAATAACATTCGT 29040
- Y Y C E N 479

20 TCCATGATTTGTACCCCCACACACCCCTTAACCAATAGTTGGCAGTTATAACCTCCGAACCGTTCCATGCGCTTGTATAACGACAGACTCTGATGGAATGTTCCAATTA 29160
W T N T G G W V V G E L N I T P V N Y G G F R E M R R T Y R V S E S P I T G I L 439

ACGTATATGCCCATATGCAAGGATAATGTTGGGAAGTCCCGAAATCGCGGCTCATGATACATCGCTGTCTAGCAAGTCCCAATTTACTCTGTAATTCGCAATACTAC 29280
T Y A A Y M C S L Q T P L G R F D G T W Q Y L R Q R A L N W N V G T I E G I S C 399

ATCGAGGGCTTGTGGGCTATTGGATACTGCACAGCGGCAACGCCCTTGTGTTATATGCTGGTGGGTTTGAACCCCTTCAGTCCCGGCGGCAATTTTCAGCTCGTATGCGTC 29400
R P S T P D N S L Q V L P L A R T N Y P Q H T N A V G E T G W A A N E A R I R G 359

CTAACAGGAAGCAATACCACGACCAAAATGTTGCTGTTAGTTGGCTTAATGCAAGATGCACTTACACCTTCTGTTGGCGTCTGTGTATACAAAAACCAAGAACACATGCT 29520
L L F G I G R G F C Q E N L Q S L A L H L R V G E R Q R R Q T Y V F V L F V H K 319

TCAGTCCGTCGCGGAAGATGTAATCTTGTCAACGTCGCAAAATACGAGCGGGATGTTGGCTGTGACCTGCGAGTTGAAGTTTGTCTGTACGTGACGCTTCTTGGGACCTT 29640
L G D A S L H L D K D V D W F V C A P I N A T V R R T S T K D T R A A E Q P G K 279

TGGCCACGCGGTTATATGCAATAATTAATCTGAATGGTATATCCAGCAGGACCAAAAAAATTAATAATGATGGAATACATGACATTGTACATCGACGTAACACTCG 29760
A V A T I N C L N D Q I T Y E L L S G F F S I F R H P F V H C Q V M A R L C E S 239

AAAACCTTATGAGCGGTTTCCATACGACGATCCATAGGAGAAACAAATGCTGTTCTGTTGGCATCCGCTGCTTATCCGATATTTCTTCTGCCGCGATGCGGATGAAC 29880
F R I L R T E M R S C G Y A S V I A T R N A D A A Q K D T Y E E A R C A A I F S 199

TAAATGAGCTTACATGCTAAGCCCCACCTTCTCCAAAGCTCAAGGAGCGTGCAGGCAATTGAATAGGTTTCTGAACCTCTAGTAGTACATCGGGTCACGTCAGCTGTG 30000
L S T V Y A R L G G G E G V T M P A T C A N F L N R L G E L L V D P D R G A Q T 159

TAAGTATTAGCTTCCAATCATGTGAGATGATGACGAAGATTGAAGCAGATTGACCCAGCATGCTCAATGTCGAGCAAAAAAATCGGTTAATGACACTTGTGGATTAGCTGTG 30120
L T N A E G I M D S P H R L I L V I S G A H E I D P R F F D T L S V Q Q I L Q T 119

TCGTTGATTTAAATYATTAACGGAGTCTAATGGTAACCTGCGGGTACCAATGAAGTTGGATTATTTGAATGTTGTCATACGATTAATAACAATTGAACGGGGGTTACTTGA 30240
T S K F N N L P L R I T V Q P N G I S T P N I Q I N N M R N I V I S R P T V Q I 79

TAGACGGGTTTCTGTACGTTTGGTGGTACATGTATCGGTTGTTGTTGACAGCTCCAAAGCGAGGGCAATTGTTAAATCGCGACTCCAAATTCGGAAGAAGCCCGAGCATAAGTCA 30360
S A T E T R K P P V H I P Q K N L G G F R P G I T L D R S W N G F F G P A Y T M 39

TATGAAGCCCGTCCCTATTTGAATAAACGGTATTTCTTAAAAGAGCTGATATTAGTTCACATAGCGTTTGTGCGTTTAAAGTAAATCGAGTGGTGGTGGTACCTCCCATAGCTG 30480
H L G N G I Q I F R N W G L L S I N T G C L T Q E N L T F H S N T P Q S G M 1

AGGGGTTAAATTCACACAATGCAATCGTGACGTGGTACTATCTGAAATGTGCTGGGGTATGTGTACACATTATACAGTCGTAGTACCGTTTATATAATGTAGGTAGGAGGAGCTAT 30600

AAAAATATTTGATTGGCGTTAAAGGTTCTTCAACTACCGTGACGTCTTTTATTAACATGCGTTTATTTAGTGTACATTATGTCTTTTCATTCGCGAGCGATGAGCTTTTTC 30720

21 M E E P I C Y D T Q K L L D D L S N L K V Q E A D N E 27
ATATCACGTTATAAAGTTAAGTCAGCTAGAAATATACCATGGAAGAACCAATTTGTTATGATACACAAAACCTTTTGATGATTAAAGTAACGTGAAGATGACGAAGCGGACACGAAA 30840

R P W S P E K T E I A R V K V V K F L R S T Q K I P A K H F I Q I W E P L H S M 67
GACCATGGTCACAGAGAAAACAGAAATCGCCAGAGTTAAGGTAGTTAAGTTTACGATCTACCCAGAAAATTCAGCTAAACATTTTATTCAGATAGGGAACCCCTGCATTCTAATA 30960

I C F V Y S N T F L A E A A F T A E N L P G L L F W R L D L D W T I E E P G N S 107
TCGTGTTTGTATATTCATACATATTTTGGCGAGGCTGCTTTCAGGCGGAAAATTTACCGGAGCTGTGTTTGGAGACTAGATCTAGACTGGAGCATAGAGGAGCCAGGATAGCT 31080

L K I L T Q L S S V V Q D S E T L H R L S A N K L R T S S K F G P V S I H F I I 147
TAAAAATTTAAACCCAGCTATCAAGTGTAGTACAAGATTCGAGAGCTTACATCGTTTACGCGCAATAAATACGAACCTCGTCTAAATTTGGACCGGTTTCGATAGACTTCATTATATA 31200

T O W I N M Y E V A L K D A T T A I E S P F T H A R I G W L E S A I A A L T Q H 187
CGGACTGGATAAATGTATGACGAGTCCCTTAAAGATGCAACACAGCATTGAATCACCATTCTACGCTGCTATTGGAATGTGGAAAGCGCCATTGAGCTTAAACACAACATA 31320

K F A I I Y D M P F V Q E G I R V L T Q Y A G W L L P F N V M W N Q I Q N S S L 227
AATTGGCATCATTTACGATATGCCATTTGTTCAAGAGGGGATTCGTGTTTAAACAAATATGACAGGATGGCTTCTCCGTTAATGTTATGTTGAATCAGATTCAAAATAGCTCACTCA 31440

T P L T R A L F I I C M I D E Y L T E T P V H S I S E L F A D T V N L I K D E A 267
CTCCCTAACACGAGCCCTTTTATAATCTGTATGATTGATGATATCTCAGGAACGCCAGTACATAGCATATCAGAAATTTTGCAGATCTGTAATTTAATTAAGATGAGGCGT 31560

F V S I E E A V T N P R T V H E S R I S S A L A Y R D P Y V F E T S P G M L A R 307
TCGTATCCATCGAAGAAGCGGTAAAGATCCAGAACCGGTGACGAGTACGAAATTCCTCAGCTCTGGCTTATCGAGACCTTATGTTTGGACATCCCGGGAATGCTTGTCTAGGA 31680

R L R L D N G I W E S N L L S L S T P G I H I E A L L H L L N S D P E A E T T S 347
GACTTAGATTAGACAATGGTATATGGGAAGCAACCTTATCTGTTGCTACCCCGGAATTCATATTGAGGCGCTGTTACATTTACTAACTCCGACCCGGAAGCGGAACCAACATCTG 31800

G S N V A E H T R G I W E K V Q A S T S P S M L I S T L A E S G F T R F S C K L 387
GAAGTATGTAGCAGAACACACCGTGGCATTGGGAAAAGGTTGAGGCTAGTACATCGCTAGTATGTTAATAGCACCTTGGCGGAATCGGGTTTACAGATTTCATGCAAAATGTC 31920

L R R F I A H H T L A G F I H G S V V A D E H I T D F Q Q T L G C L A L V G G L 427
TACGTGGTTATTTGCTCACCACACTCGCGGTTTATTCACGGAAGCGTGTAGCAGAGGATATTACAGATTTCACAAACACTAGGATGCTCGCTTATGTTGGTGGACTGG 32040

A Y Q L V E T Y A P T T E Y V L T Y T R T V N E T E K R Y E T L L P A L G L P P 467
CATACCAATTAGTGGAAAGTACGCTCTTACACGAGTATGTTTACATATACACGGAAGTAAACGAGACGGAAGACGATGAAACGCTATTACCCGCTTAGGATTACCAACCGG 32160

G G L G Q I M R R C F A P R P L I E S I Q A T R V I L L N E I S H A E A R E T T 507
GAGGCTGGGACAAATATGCGGCGCTGTTTGTCCAGACCCCTTATTGAAAGTATACAAGGACACGCGTAATACCTACTTAAAGAAATTTACAGAAAGTATAGAGAGACAACAT 32280

Y F K Q T H N Q S S G A L L P Q A G Q S A V R E A V L T W F D L R M D S R W G I 547
ATTTAAGCAACACATAATCAATCCTCAGGTGCGTTATTACCAAGCAGGACAAAGTCCGCTGACGGAAGCCCTACTAACCTGGTTGAGTACGATGGAATGAGGATGATTA 32400

T P P V D V G M T P P I C V D P P A T G L E A V M I T E A L K I A Y P T E Y N R 587
CTCCCCGGTGGATGTTGATGACACTCCTATTGTTGTTGATCCACGCGTACAGGTTGGAAGCTGTATGATAACAGAGCACTAAAGATTGATATCTACCGAATATAATCGCT 32520

S S V F V E P S F V P Y I I A T S T L D A L S A T I A L S F D T R G I Q Q A L S 627
CTAGCGTGTGTTGGAAACGCTGTTTGTGCTTATATTATTGCAACAGCACGCTTGATGCCCTTCGGCAACATAGCTTTGTCTTTGATACAGGGGATACAGCAAGCCTTGTCTA 32640

I L Q W A R D Y G S G T V P N A D G Y R T K L S A L I T I L E P F T R T H P P V 667
TTCCTCAGTGGCTCGGATTTAGGATCCGGAACCGTGCCCAATGAGATGGATATCGCACAAAATCTGCTTATAACAATATTAGAACCTTTTACCCGTACACACCCCAAGTAC 32760

L L P S H V S T I D S L I C E L H R T V G I A V D L L P Q H V R P L V P D R P S 707
TATTACATCTCAGCTTCTACTATAGATTCCCTTATATGCGAATCTCATCGACTGTGGCATTTGCGATTGCGCTTCCCAAGCAGTCCGCTTGTGGTCTTGACCGTCTTCTA 32880

I T N S V F L A T L Y Y D E L Y G R W T R L D K T S Q A L V E N F T S N A L V Y 747
TTACAAATAGCGTTTTTATAGCACTCTTATTATGATGAATCTTACGGCTGTGGACCCGACTGGATAAAACATCGCAGGCTTGGTTGAAAATTTTACATCCACGCGTTAGTGGTTT 33000

S R Y M L M L Q K F F A C R F Y P T P O L Q A V G I C N P K V E R D E Q F G V W 787
CTCGGTACATGTTAATGTTACAAAATTTTTGCGTGTGTTTTATCCAAGCCAGATCTCAGGCTGTTGGTATCTGTAAACCAAGGTTGAACGCGATGAACATTTGGGGATGCG 33120

R L N D L A D A V G H I V G T I O G I R T Q M R V G I S S L R T I M A D A S S A 827
GTTTAAACGATCTTGTGATGCGGTTGGTCAATTGTTGGGACAATACAGGAATCCGAACGCAAAATGAGAGTGGGAATATCCAGCTGCGCACAAATTTAGCGGATGCTTCTCAGCCC 33240

L R E C E N L M T K T S T S A I G P L F S T M A S R Y A R F T Q D Q N D I L M R 867
TTAGGGAATGTGAAAATTTAATGACTAAACCTCCACTTCTGCTATTGGGCTCTTTTTTCAACGATGGCTTCCGGTATGACAGGTTTACACAGGATCAATGGAATTTAATGCTGT 33360

VZV DNA sequence

1773

V D K L T T G E N I P G L A N V E I F L N R W E R I A T A C R H A T A V P S A E 907
T T G A C A A C T A A C A C G G A G A A A T A C C G G T C T G C A A A T G T A G A G A T T T T T T A A A T A G G T G G G A C G A A T A G C A A C A G C T T G T A G G C A T G C C A C G C A G T C C C G T C G G C C A A T 33480

S I A T V C N E L R R G L K N I Q E D R V N A P T S Y M S H A R N L E O H K A A 947
C T A T T G C A A C C G T G T A A T G A A T T G A G G C G G T T T A A A A A T A T A C A A G A G A T G T G T A A T G C C C A A C C T C A T A T A G A T C A C G C C G A A A T C T G A A G A T C A C A A G C A G C A G 33600

V S F V M D S R Q Q F I V D S G P Q M G A V L T S Q C N I G T W E N V N A T F L 987
T T T C A T T C G T A T G G A C T C C A G G C A A C A G T T A T T G G A T T C T G G A C C T C A G A T G G G C G C G T T T A A C T T C A C A T G A A T A T A G G A C A T G G G A A T G A A A T G C A A C G T T T T T A C 33720

H O N V K I T T T V R D V I S E A P T L I I G Q R W L R P D E I L S N V D L R L 1027
A T G A T A A T G T T A A A A T A A C G A C A C G G T C A G A G C G T A A T T T C A G A G G C T C C G A C G T G A T A A T A G G A C A A G A T G G C T T C G T C C A G A T G A G A T T T A T C T A A T G T A G A T T T G C G T C T G 33840

G V P G N T S G S D P - 1038
G C G T A C C C G G A A T A C A A G T G G A G T G A C C C T A A T A T A A A C A G G C G T G T T A T G T A C A T T A A G T A T T G T G G T T T T A T T G A C T G G G C G T T C G T T G T A A C G C T G T T G T G C T A 33960

G T A T T T T C A A C C T C C T A G G T T T T G G A G C T A C A C G T C T T A T C A A C G C T C T T T G G G A T T T G A A T C A T C G T A A C G T A G C G T C C C T A C C A G T T G A G C G C T A A T T T C G T A A G C A A T A 34080

22 M D I I P P I A V T V A G V G S R N Q F D G A L G P A S G L S C L R T S L S F 39
A A A T G G A T A T A A T C C C C T A T A G C T G T C A C T G T T G C G G A G T G G G A A G C G T A A T C A A T T T G A C G G T G C C T G G G A C C G G C T C A G G T C T G C A T G T T A A G A C A T C T T T A T C G T T T 34200

L H M T Y A H G I N A T L S S D M I D G C L Q E G A A W T T D L S N M G R G V P 79
T G C A T A T G A C A T A G C G A T G G A A T A A T G C A A C C T G T C A T C A G A C A T G A T T G A T G G A T G T T A C A A G A G G G T C A G C A T G G A C T A C G G A T C T G T A A T A T G G G A G G G G T G C C C A G 34320

D M C A L V D L P N R I S Y I K L G D T T S T C C V L S R I Y G D S H F F T V P 119
A T A T G T G C T C T T G T A T C T C C C A C T G A A T T C A T A T A T A A C T G G G G A C A C T A C C A G T A C G T G C G G T T T T G T C T A G A A T A C G G C A T A G C C A T T T T T T A C C G T C C A G 34440

D E G F M C T Q I P A R A F F D D V M G R E E S Y T I I T V D S T G W A I Y R 159
A C G A G G G T T T A T G T G C A C A A A T C C C G C T A G A G C G T T T T C G A T G A T G T G G A T G G G A C G T G A A G A G C G T A C A A T T A A C T G T A G A C T A C C G G A A T G G C C A T C A T C G T C 34560

Q G N I S F I F D P H G H G T I G Q A V V V R V N T T D V Y S V I A S E Y T H R 199
A G G G A A C A T A T C T T T T T T T G A T C C A C A T G G C C A T G G G A C T A G G A C A G C G T A G T G T T G T G G G T A A T A C C A G G A T G T G A C T T A T A T C G C A T G G A G T A C C C A C C G C C 34680

P D N V E S Q M A A A L V F F V T A N D G P V S E E A L S S A V T L I Y G S C D 239
C C G A T A C G T A G A A T C C A A T G G G C G T C A T T A G T T T T T T G T C A C C G A A C A C G C G C C G T A A G C G A A G A A G C G T A T C T C G G C A G T A A C G C T A T A T A C G G A A G C T G T G A T A 34800

T Y F T D E Q Y C E K L V T A Q H P L L S P P N S T T I V L N K S S I V P L H 279
C A T A T T T T A C A G A T G A A C A A T T G C G A A A A C T G G T T A C A G C T A C A C T C G G T C T T C T T C A C C T C T A A T T C C A G C A A T T G T C T A A T A A T C G T C T A T A G T A C C T C T C A C C 34920

Q N V G E S V S L E A T L H S T L T N T V A L D P R C S Y S E V D P N H A V L E 319
A A A C A T T G G T G A A A G T G A T C T T G G A A G C A A C C T A C A T T C A A C G T T A A C A A C A C G G T T G C A C T G G A C C C T A G A T G T A G T T A C G C G A G G T T G A C C T T G G C A T G C G G T T C T A G A A A 35040

T T S T G S G V L D C R R R R R P S N T P P S S E E N L A C I D D G L V M N T H 359
C A A C C T C G A C T G G G T C T G G C G T T T G G A T T G C T G C T G A G A C C G C T C C T C A T G G A C T C C C T T C A A G C G A G G A A A T T T A G C T T G T A T C G A C A T G G C T T G G T A A A T A A T A C A C A T T 35160

S T D N L H K P A K K V L K F K P T V D V P D K T Q V A H V L P R L R E V A N T 399
C C A C G G A T A A T T A C A T A A C C C G T A A A A G G T T C T C A A A T T A A C C A A C T G T A G A C G T G C C G G A T A A A C A C A A G T G G C A C A T G T A T A C C C C G C T A C G A A A G T T G C T A A C A C C C 35280

P D V V L N V S N V D T P E S S P T F S R N M N V G S S L K D R K P F L F E Q S 439
C A G A C G T T G T G T A A T G A T C C A A T G A T A G A T C C C T G A A T C C A C C C A C T T T T C A C G G A A C A T G A A T G A G A A G C A G T T G A A A G A T C G G A A G C A T T T C T A T T G A A C A G A G T G 35400

G D V N M V V E K L L Q H G H E I S N G V V Q N A V G T L D T V I T G H T N V P 479
G T G A T G C A A C A T G G T T G C G A A A A C T A C T A C A A C A T G G G C A T G A A A T A G C A A T G G A T A C G T A C A A A A T G C G G T G G G T A C G T T G G A T A C T G T T A T T A C C G G C A T A C A A A T G T T C C C A 35520

I M V T R P L V M P D E K D P L E L F I N L T I L R L T G F V V E N G T R T H H 519
T T T G G T A A C A G G C C T T G G T A T G C C A G A C G A A A A G A T C C A T T G G A G C T T T T A T A A C C T C A C C A T T T T G C G T T A A C G G G A T T T G T G G T G G A A A T G G A A C A C G T A C A C A T C A T G 35640

G A T S V Y S D F I G P L G E I L T G F P S A A E L I R V T S L I L T N M P G A 559
G T G C T A C A A G C G T T G A T C A G A C T T T A G G T C C C T T G G G G A A A T T T A A C A G G A T T T C C C T C C G C G C G A A C T A T A C G C G T T A C A A G T T G A T A T T A C A A A C A T G C C G G G G C G G 35760

E Y A I K T V L R K K C T I G M L I A K F G L V A M R V Q D T Y G A L H A E L 599
A A T A T G C T A T T A A A C T G T T C C G G A A A A A T G T A C A A T T G G C A T G C T A T T A T C G T A A G T T T G G T C A G T T G C C A T G C G G G T T C A G G A T A C A A C C G G C C T T A C A T G C C G A C T A G 35880

O V L E A D L G G S S P I D L Y S R L S T G L I S I L N S P I I S H P G L F A E 639
A T G T T T A G A A C G G A C T A G G A G G T C G T G C C C A T A G A C C T C A T T C T A G A C T G T C A C A G G T C T T A A G T A T A C T A A A T T G C C T A T T A T T C T A T C C C G G A C T T T T T G C C G A G C 36000

L I P T R T G S L S E R I R L L C E L V S A R E T R Y M R E H T A L V S S V K A 679
T T A T T C C A A C C G T A C A G G G C C C T G T G A C A A A T A C G T C T T C T T G T G A A T A G T C T C G G C C G G A G A C A C G T A T A T G C G T G A A C A C C G C C T T G T T C T A G T A A A G C C T 36120

L E N A L R S T R N K I D A I Q I P E V P Q E P P E E T D I P P E E L I R R V Y 719
T A G A G A A T G C A T T A C C G T C A C C G C A A A A A T T G A T G C C A T T C A A A T A C C A G A A G T T C C C C A G A A C C C C G A A A A C C G A C A T T C C A C C G A A G A G T T A A T C G G C G T G T A T A G 36240

E I R S E V T M L L T S A V T E Y F T R G V L V S T R A L I A E Q S P R R F R V 759
A G A T A C A T C C G A A G T A C A A T G C T A T T G A C C T G G C T G T T A C A G A A T A C T C A C C C G G A G T G T T A T A G C A C A C G G C C T T G A T C G T G A A C A A T C C C C T A G G C G T T T T C G G G T C G 36360

A T A S T A P I Q R L L D S L P E F O A K L T A I I S S L S I N H P P P E T I Q N	799
CGACCGCAAGTACGGACCCATTCAACGGCTTTTAGATCTCTCCGGAATTCGACGCTAAATTAACGGCAATCATATCGTCCCTGCTATACCCCTCCCTCGAGACTATACAAATC	36480
L P V V S L L K E L I K E G E D L N T D T A L V S W L S V V G E A Q T A G Y L S	839
YCCCCGTGCTATCTGTGTTAAAGAGCTTATTAAGAAGGGGAAGATTAAACACAGACACGGCTCTGATCTGTTGTTATCTGATGTCGGGGAAGCTCAACCGCAGGTACTTATCCA	36600
R R E F D E L S R T I K T I N T R A T Q R A S A E A E L S C F M T L S A A V D Q	879
GACGAGATTCGATGAATTATCAGCTACAATTAACCAATTAATACACGGCAACGCAACGGGCTTCCGCGGAAGCAGAGTTGTTGCTTTAATACGCTAAGCGCGGCGTACGACCAAG	36720
A V K D Y E T Y N N G : E V K Y P E I T R D D L L A T I V R A T D D L V R O I K I	919
CCGTAAGGACTATGAACATATAACAATGGTGAAGTCAAGTATCTGAAATAACACGGGATGATTTATTAGCAACAATTGTACGCTACAGAGATTGGTGCAGAGATAAAATTT	36840
L S D P M I Q S G L Q P S I K R R L E T R L K E V Q T Y A N E A R T T Q D T I K	959
TAAGTATCAATGATCAATCCGGTTTCAACCTTCGATTAAGACAGTGGAAACAGGCTTAAGAGGTTACAGAGTATGCAACAGGAGCCCAACACAGGACCAATAAAGA	36960
S R K Q A A Y N K L G G L L R P V T G F V G L R A A V D L L P E L A S E L D V Q	999
GTCGAAACAGGCGGATATAATAAATCAGGGGTTACTTCGCCCGGTAAACGGTTTGTGGGACTTAGGGCTGCAAGTAGATTATTACCGGAATCTGCTGAGTATAGATGCAAG	37080
G A L V N L R T K V L E A P V E I R S Q L T G D F W A L F N Q Y R D I L E H P G	1039
GAGCCCTGGTAATCTCAGGACCAAGCTTAGAGGCGCGGTAGAGATCCGTTCTCACTACGGGTGATTTCTGGGCGTTATTAAACCAATTCAGAGACTTTTGAACATCCCGAA	37200
N A R T S V L G G L G A C F T A I I E I V P I P T E Y R P S L L A F F G D V A D	1079
ACGACGCAATCTGTCTAGGAGGACTGGGAGCTTGTTTACAGCTATTATCGAAATGTGCGGATACCTACGGAGTATAGACATCAATGCTTGGCTTTTGGTGACGTGGCAGATG	37320
V L A S D I A T V S T N P E S E S A I N A V V A T L S K A T L V S S T Y P A L S	1119
TGCTTGATCCGACATCGGACCGATCTACTAACCCGGAAGTGAGTCCGCCATAAACGCTGTTGTTGCAACTCTTAGTAAAGCGACGTTAGTTTCTATCAGAGTCCAGCTTATCT	37440
F V L S L Y K K Y Q A L Q Q E I T N T H K L T E L O K Q L G D D F S T L A V S S	1159
TGTGTTGCTTATATAAAAAATATCAGGCTTACAAACGAAGTACGAATACCCATAAGTGTGACTGAATTAACAAACAACTTGGAGATGACTTCCACCTAGCTGTCTCATCTG	37560
G H L K F I S S S N Y D D Y E I N D A I L S I Q T N V H A L M D T V K L V E V E	1199
GACACTGAAGTTTATATCACTTCAATGTAGATGATTATGAATAAAGGATGCGATATTATCAATACAAACAAAGTGACGCGCTTAATGGATACGGTTAACTTGTGAGTTGAAC	37680
L Q K L P P H C I A G T S T L S R V V K D L H K L V T M A H E K K E Q A K V L I	1239
TGCAAAAGCTACCCCCCATTTGTTTGTGCTGGGACATCATCTTATCTCGAGTAGTAAGGATCTTCAAACTCGTCAATGGCAGATGAGAAGGAAGGACAGGAAAGTGTATTA	37800
T D C E R A H K Q Q T T R V L Y E R W T R D I I A C L E A M E T R H I F N G T E	1279
CCGATTGTGAACGTGCACATAAACACAAACGACTCGGGTTTGTATGAGCGTTGGACAGTGATATTATAGCATGTGCGAGGCAATGGAAACCGCCATATATTAAACGGGACAGAAC	37920
L A R L R D M A A A G G F D I H A V Y P Q A R Q V V A A C E T T A V T A L D T V	1319
TGGACGCTTGGCAGATATGGCCGCTGCGGGAGGGTTGATATACACGAGTTTACCCACAAGCAGTCAAGTTGTAGCGCATGTGAACACAGCCGTTACGGATATAGATATCTGT	38040
F R H N P Y T P E N T N I P P L A L L R G L T W F D D F S I T A P V F T V M F	1359
YTCGCCAATCAATATACCCCGAAATACAAATATCCACACCTTTGGCTTGTGAAGAGGTTAACTGGTTGATGATTTCGATTACGGCTCCGATTACCGTTATGTTTC	38160
P G V S I E G L L L L M R I R A V V L L S A D T S I N G I P N Y R D M I L R T S	1399
CAGGTGTAGTATGAGGACTCCTTCTGCTTATGCTATTGCGCGGTTGTGTTATTCGCGGATACGCTTATTAATGGAATACCTAACTACCGAGATATGATATACGAACCTCGG	38280
G D L L Q I P A L A G Y V D F Y T R S Y D Q F I T E S V T L S E L R A D I R Q A	1439
GGGATCTATTACAAATACCCGACTTGGCTGGGTATGTTGATTTTACACAGGCTTATGATCAGTTTATAACCGAAAGTGAACGTTAAGTGAATATAGAGAGACATCAGACAGGCTG	38400
A G A K L T E A N K A L E E V T H V R A H E T A K L A L K E G V F I T L P S E G	1479
CCGGGCTAAACTTACAGAAAGCAATAAGGCTTGGAGGAAGTAACATGTTCCGGCACACGAAACGGCTAAACTTGCATTAAGAAGGTGCTTCTTACATTACCAAGCGAAGGTT	38520
L L I R A I E Y F T T F D H K R F I G T A Y E R V L Q T W V D R D L K E A N A E	1519
TATTGATTCGGGCTATAGAGTATTTTACAACTTCGATCATAAACGATTTATAGGAACGGCATATGAAGAGTTTACAAACATGGTAGACCGGATCTAAAGGAGGCCAACGACAGGC	38640
L A Q F R M V C Q A T K N R A I Q I L Q N I V D T A N A T E Q Q E D V D F T N L	1559
TTCACAGTTTCGATGTTGTGTCAGGCAACAAAGAACGTCGAATACAAATTTACAAACATTTGTTGATACGGCAATGCCACTGAGCAACAAGAGACGTGGATTCTACTAACCTGA	38760
K T L L K L T P P P K T I A L A I D R S T S V Q D I V T Q F A L L L G R L E E E	1599
AGACGTTATTAACAACTACCCCCCTCCCAAAACAAATGCAATGGCCATGTAGATCTACTTCGCTCAGGACATTGTCACGAGTTGCTTGTAGGGGCTTGGAGGAAGAAA	38880
T G T L D I Q A V D N M Y Q A R N I I D S H P L S V R I D G T G P L H T Y K D R	1639
CTGGTACGTTGGACATTCAGGCGGTGACTGGATGTACCAAGCTCGCAATATTATGACTCCCACTCAAGTGCGTATAGAGGATACCGGCCCCCTGCTACTATATAAGATAGGG	39000
V D K L Y A L R T K L D L L R R R I E T G E V T N D D A W T T F K R E T G D M L	1679
TGGATAAATTTATGCGTTACGAACATAATAGATCTCTACGACGACGAATAGAACCGGTGAGGTTACGTGGGACGATGATGGACAACATTTAAAGAGAAACGGGGATATGTTGG	39120
A S G D T Y A T S V D S I K A L Q A S A S V V D M L C S E P E F F L L P V E T K	1719
CATCGGGGACACGTACGCTACTTCGATAGATATAGAGGCACTCCAGGATCGGCGTCTGTGGTTGACATGCTTGTTCGAACCGAATTTTTTTTATGCTGTGGAAACGAAAA	39240
N R L Q K K Q Q E R K T A L D V V L Q K Q R Q F E E T A S R L R A L I E R I P T	1759
ACCGCTTCCAAAAAAGCAACAGGAACGTAAACCGGCTTGGATGTTGTTGCGAAAAACAAAGACAGTTTGAAGAGACCGGCTCTGCTTACGAGCTTAATGAACGATTTTCAACGG	39360

ESDHDVLRMLLRDFDQFTHLPINIKTQYMTFRNLLMVRLG 1799
AGAGTGACCATGCGTTCTCGTATGTTACGTGATTGCGATCAATTTACACATTTCCTATATGGATAAAACACAGTATGACATTTGCAATTTACTCATGGTACGGTTAGGCT 39480

LYASYAEIFPPASPNGVFAPIPAMSGVCLLEDQSR CIRARV 1839
TGTATGCAAGTTATGCTGAGATTTTCCACCCGCGCTCCAAACGGAGTATTGCTCTTATCCCGCCATGTCCGGGTATGCTAGAGACCAATCCCGATGCATTCGCGCGCGGGTGG 39600

AAFWMGEASVYQTFREARSSIDALFGKNLTFYLDTDGVPLR 1879
CCCGTTTATGGGGAGGGCTGCTGGTGCAACGTTTAGGGAAGCCAGATCTCTATAGACGCTTGTGTTGGAAAAATTTAACCTTTTACTTGGATAGTGGGGTCCACTTCGAT 39720

YRVCKYSVGVKLGTM LSCQGGLSLRPALPDEGIVEETTLS 1919
ATAGAGTGTGTATAATCAGTTGGGGTAACTTGAACCATGCTATGACGTAGGGTGGATTATCTTACACCGGCACTTCCGATGAAGGATTGTGGAAGAACTACACTATCGG 39840

ALRVANEVNEELRIEYESAIKSGFSAFSTFVRHR'HAENWGKT 1959
CATTACGCGTGGCAATGAGGTCAATGAGCTACGCAATGAATACGAATCCGCTATAAAATCCGGGTTTCTGCCCTTTTCCACCTTGTGTTAGGCATGCCACGCCGAATGGGGTAAACCA 39960

NARRAIAEIIYAGLITTTTLTRQYGVHWDKLIYSFEKHHLTS 1999
ACGACGAGAGCCATTCAGAGATATACGCCGCGCTTAAACAAACATTGACACGACATACGGGGTTCATTGGGACAAGCTTATTTATCTTTTGGAAAAACCACTTAACCTTCG 40080

VMGNGLTKPIQRRGGOVRVLELTLSDIVTILVATTPVHLLN 2039
TAATGGCAATGGCACTAACTAAACCAATCCAGAGAGGGGTGATGACGCGTATAGAGTTAACCTATCTGATATTGTAATTTTGGTGGCACAACCCCGGTACATCTCTCAATT 40200

FARLDLIKQHEYMARTLRPVIEAAFRGRLLVRS LDGDPKG 2079
TTGCTAGATTGGATTAATAACACAGATAGTATATGGCCGTACCCCTCAGACCGGCAATGACGGCCGCAATTAAGGTCGTTTACTGCTGCTCATTGGATGGAGACCCGAAAGGCA 40320

NARAFFNAAPS KHKLP LALG SNQDPTGGRIFAFRMADWKL 2119
ATGCCCGGGCTTTTAAATCCGCGCCCTTCAACATAAACTCCCGTTAGCTTGGATCAAAACAGATCTTACC GGCGGGAATATTGCAATTCGGATGGCAGATTGGAACTTG 40440

VKNPQKITDPPAPMQLSPPPGVKANVDAVTRIMATDRLAT 2159
TAAATGCAACAGAAAAATACGAGCTTTTGGCCATGGCAACTTCCCCCCCCCGGGGTAAGGCCAATGTCGATGCAAGTACCCTATAATGGCAACAGATCGCTTGGCAACCA 40560

ITVLGRMCLPPIISLVSNWNTLOPEEFAYRTQDDVDIIVDA 2199
TYACTGTACTTGGCGCATGTGCTCCCGCAATTTCTTATGTCATGTAATGGAATACGCTGCAACCGAGGAATTCGCATACAGAACACAAAGATGATGGACATTATAGTTGATGCGA 40680

RLDLSSTLNARFDTAPSNTTLEWNTDRKVITDAYIQTGAT 2239
GACTGGATTTGCTACACGCTTAATGCAAGATTGATACCGCTCCCGCAATACCACTGATAGATGGAATACAGACCGTAAGTAATACAGATGCTTATATTCAACCGGGCAACGA 40800

YVFTVTGAAPTHVSNVTAFDIATTA ILFGAPLV IAMELTS 2279
CAGTTTTACAGTAACGGGGCGGCAACCTACGTTCTAATGTAACAGCGTTTGACATAGCACTACGCTATTTTATTTGGGGCTCTTTGGTTATGCAATGGAACTTACATCGG 40920

VFSQNSGLTLGLKLFD SRHMATDSG ISSAVSPDI VSWGLR 2319
TTTTTTCACAAAATCCGGACTTCTTTGGGGTAAATATTGCAATCCCGCATATGGCTACAGATTCGGGTATATCTCAGCGTATCTCCGATATTGTTCTTGGGGGTACGTT 41040

LLHMDPHPIENACLI VQLEKLSALIANKPLTNWPPCLLLL 2359
TACTGATATGGATCCACCAATGAAATGCAATGTTAATGTCACATAGAAAACCTGTCGGCTCATGCAACAAACCTCTTACAAACATCCCCGTGTTTACTGCTATTGG 41160

DEHMNPSYVLWERKDSIPAPDYVVFVWGPESLIDLPIYIDS 2399
ACGAACATGATGATCCCTTATGTTTATGGGAACGAAAAGACTCGATTCCAGCTCCGATTATGTTGCTTTTGGGGGCCAGAACTCTTATGATTGCGTACATCGACTCCGATG 41280

EDSFPSCPDDPFYSQIIAGYAPQGPPLNDITDFYPTPLF 2439
AGGACTTTTCCCTCGTGTCCCGATGATCCATTTTACTGCAAAATATTGCGGTTATGCGGCCCAAGGCCCCCAACCTCGACACAACTGATTTTACCCAAACGGAGCACTATTTA 41400

KSPVQVVRSSKCKKMPVYRPAQPAQPAQPAQPAQPAQPAQ 2479
AGTCTCCGTTCAAGTTGTAGAGATTCCAAATGTAATAAATGCCCCGCGCGAGCCCGCGAGCCCGCGAGCCCGCGAGACCGTCCAGCCCGCGAGCCCA 41520
< A > < A > < A > < A > < A > < B > < A > < I >
<----- reiteration R3 ----->

IEPGTQIVVQNFKKPQSVKTTLSQKDIPLYVETESETAVL 2519
TAGAACCGGACACAAATAGTGGTACAAAATTTAAGAAACCCCAAGCGTAAAAACAACCTTAGCCAAAAAGATATTCCCTTGTATGTGGAAACCGAATCAGAAACGGCTGTGCTTA 41640

IPKQLTTSIKTTVCKSITPPNNQLSDWKNNPQONQTLNQA 2559
TACCTAAGCAATTAACCACTCCATTAATAACACCGCTTGTAAAGATTATCCCAACCAATAACCAATTTGCGGATTGGAAAAATATCCACAGCAAAACCAACGTTAAACCAAGCGT 41760

FSKPILEITSIPTDSSISYRTWIEKSNQTQKR'HNQNDPRMY 2599
TCAGTAACCAACTACTGAGATTACCTCATTCGACAGATGACTCGATCTTACCGGACTTGGATTGAAAAATCAATGAACACAAAACCGCATCAAAATGACCTCGAATGTATA 41880

NSKT V FHPVNNQLPSMVDYAADAPQTDLLTNYKTRQPSN 2639
ACTCCAAAACAGTATCCACCTGTAATAACCAATTAACCTTCTGGGTGACACGGCAGCGATGCCCCCAACGAGCTATTGACAACTATAAAACAGACAGCGCTCGCCAACT 42000

FPRDVHTWGVSSNPFNSPNRDLVQSD FSEPSDGYSSSEEN 2679
TTCCGGCGGACGTACACATAGGGGCTATCTTCAACCCGTTAACTACCGAACAGACAGCTATATCAAGTGATTTAGTGAACTTCTGACGGCTATAGCAGTGAGAGTGAAAAAT 42120

SIVLSLDEHRSCRVP R HV RVVNADVVTGRRYVRGTALGAL 2719
CTATCGTACTAAGTCTCGACGAACATCGGTATGTCGCTTCTAGGACGTCAGCGTGTGTAATGCGGATGTAGTACC GGTCAGCTTATGCCAGGGACCGCTTGGGAGCACTGG 42240

ALLSQACRRMIDNVRYTRKLLMDHTE DIFQGLGYVKLLLD 2759
CACTGTTAAGCCAGGATGTGGCGTATGATGACACGTTAGATATACCGTAACTTTTATGAGCACACGGAAGATATTTCAAGGCTGGGGTATGTTAAATGTTATTAGATG 42360

VZV DNA sequence

1777

2763
12480
220
285
45360
325
45480
365
45600
405
45720
445
45840
485
45960
525
46080
27
25
565
46200
65
585
46320
105
46440
145
46560
185
46680
225
46800
265
46920
305
47040
333
47160
1160
47280
1120
47400
1080
47520
1040
47640
1000
47760
960
47880
920
48000
880
48120
840

R K S F Y F N Q D G L T G G P Q P L A A A L A N W K D C A R M V D C S S S E H R
C G T A A G T C A T T T A C T T A A T C A G G A C G A C T G A C T G G A G G C C T C A A C C T T T A G C G G C C G C T T G G C T A A T T G G A A A G A T T G C G C G C G A T G G T G A C T G T C A T C A T C G G A A C A T C G C
T S G M I T C A E R A L K E D I E F E D I L I D K L K K S S Y V E A A W G Y A D
A C A A G T G G G A T G A T T A C C T G C G G A A C G T G C A T T A A A G A G G A T A T A G A G T T T G A A G A T A T A T A A T A G A C A A A C T T A A A A A T C G T T A C G T A G A A G C A G C T T G G G T T A C G C A G A C
L A L L L L S G V A T M W N V D E R T N C A I E T R V G C V K S Y M Q A W R I E N
T T G C T T T A T T A C T G A G T G G G T T G C T A C T T G G A A T G T A G A C G A C G T A C A A A T T G T C T A T A G A A A C T C G C G T T G G A T G T T A A A T C A T A C T G C A G G C A A C C G A T T G A A A C
S R D V P K Q F S K F T S E D A C P E V A F G P I L L T T L K N A K C R G R T N
T C C A G G A C G T T C C A A A C A A T T T C C A A A T T A C G A C G A G G A T G C C T G T C C G A A G T A G C A T T T G G G C C T A T T T G T T A A C T A C C T T A A A A A C G A A A G T C C G T G G T C G C A C G A A T
T E C N L C C L L T I G H Y W I A L R Q F K R D I L A Y S A N N T S L F D C I E
A C C G A T G C A T G T T A T G T T T A A A C C A T G G G A C T A T T G G A T C G C T T T G C G G A G T T A A A A G G A T A T A T A G C A T A C T C A G C A A A T A A C A A G T T T A T T G A C T G T A C G A A
P V I N A W S L D M P I K L K F P F N D E G R F I T I V K A A G S E A V Y K H L
C C T G A T C A A T G A T G G A C C T A G A T A A C C C A T A A A C T A A A T T C C A T T A A T G A T A G A G G T C G A T T A A C C A T T G T A A A A G C A G A G G T T C C G A G G C C G T A T A A A C A T T T A
F C D L L C A L S E L Q T N P K I L F A H P T T A D K E V L E L Y K A Q L A A Q
T T T T G C A T C C T A T G C C T C T C G G A A T T A C A G A C A A C C T A A A A T T T A T T G C C C A C C T A C A A C C G G A T A A G G A A G T T G G A G T T A T A A A G C C C A C T G G C G C A C A A
M H L K P T R F F H A N Q P P M P H S Y E M E D L
N R F E G R V C A G L W T L A Y A F K A Y Q I F P R K P T A N A A F I R D G G L
A A C A G A T T T G A A G T C G T A T G T G C T G G C C T G T G G A C A T T G G C G T A T G C A T T A A A G C C T A C C A G A T T T T C C A C G A A A C C A A C C C C A T G C C G A T T C A T A C G A G A T G G A G G A C T T
C F D D M Q Y R W S P S N T P Y R S M S R R Y K S V S R S G P S M R V R S R T P
M L R R H A I S L V S L E H T L S K Y V -
A T G C T C G A C A C A T G C A A T T C G C T C C C T G A A C A C A C C C T A T C G A A G T A T G T C A G G C A T A A A T C C G T A T C T C G A G C G G C C T T C G A T G C G T A C G C T C C A G A C G C C
C R R Q T I R G K L M S K E R S V Y R H Y F M Y I A R S P P E E L A T V R G L I
A T G C G C C G T C A A C A C T T G A G G A A A A C T A T G T C A A A G A G C G G C T G T G T A C C G C A T T A T T T A A T T A C T C G A A G G T C C C C C C A G A A G A C T A G C T A C C G T A G A G G C T T A A T
V P I I K T T P Y T L P F N L G Q T V A D N C L S L S G M G Y H L G L G G Y C P
C G T G C A A T T A T A A G A C A C C C T G C A C C T T C C G T T A A C T T G G G T C A G A C A G T G G C G G A T A A C T G C C G T A T C C G A A T G G G T A T C A T T A A G T C T C G A G G T A T T G T C C
T C T A S G E P R L C R T D R A A L I L A Y V Q Q L N N I Y E Y R V F L A S I L
G A C A T G C A C T C G A T C G A G A A C C G C T A T G T C G A A C G A T C G G C G C C T C G A T A C T A G C A T A T G T T C A G C A G C T T A A C A A C A T A C G A A T A C G T G T G T T T C T T G C A T C C A T T T
A L S D R A N W Q A A S A E P L L S S V L A Q P E L F F M Y H I M R E G G M R D
G G C G T A T C A G A C G A C C A A C A T G C A A G C A G C G C C T G A A C C C T A T T G T C G A G C G T A T T G C A C A A C C G G A A T T A T T T T A T G T A T A T T A T G A G G A G G G G G C A T G C G A G A
I R V L F Y R D G D A G G F M N Y V I F P G K S V H L H Y R L I D H I Q A A C R
T A T A C G C G A C T T T T A T C G T G A T G G A G A T G C C G A G G G T T A T G A T G T A T A T A T T C C G G G A A A C T G T T C A C C T C C A T T A C A G A C T A A T C A T A T A C A G C C C G T G T C G
G Y K I V A H V N Q T T F L L S V C R N P E Q Q T E T V V P S I G T S D V Y C K
G G G G T A A A A T A G T C G C A C A C G T T T G C A G A C A A C A A C A A C A G A G A C T G T G G T C C A T C C A T T G G A A C A T C G G A C G T T A C T G T A A
M C D L N F D G E L L E Y K R L Y A L F D D F V P P R -
A A T G T G T G A C C T A A C T T G A T G G A G A A T T G T T G G A A T A C A A A G A C T C A C G A T T A T T G A T G A C T T T G T C C C T C G G T G A T T C A G C T T C A G T G T C A T T T A T A T T C C C A G
- S O H L I A K P I C F V R R M I Q H S Q E E E T I E A E T N M K N D W
C A C G G G C G T G T A T A C A A A A A G C C T G C C G C T G A A G C G G T T A G C A T T T A A C G T T A A C A C T C G T C T C T G G A A T A A A C G T T T A A A A G C C G T T C T G T G A G T T A G T G C G T T
C P A N I C V F G A A Q L R N L M K V N V R T E P I F R K L L R E T L K T D N
C C A A A T A A C G C C T A A A A G T A C A C T C G C C G C C A A T G A G A T G A G A A A A A T A A T A G T C A A T G T T A A A G A C A C C C G T G T G A T T A C G T G A A T G G G A T C T C C G T A A G T C A G A T A T T
G F L A K F T V S A T G I L H S F Y Y D I N L S L G H S T V H I P D E A L D S I
A T T A A C T A C G C T T T G C T C C C C A C C G T T A C T G C G G T A T T C T G T A A A G G A T C C A C G T A G C A A A G C T A C A T T T T T G C A T C A G C C T C C A C T T C G T G T G G G G G C C A A T A A C A
I L K R K A E G C R K G A T N Q L P D G R L L A V S K A D A E V E D T P A V I V
T A A G G G A T G C G T T C G A A C G T T T G G G A T T G A C C T G T C A T T A C T A A T T A T A A T A C T G T T A A G T A G C A A G C A A C G C A A G C G G T T T A T G T A G G C G G A T G T G G A C G A C T A A G C T G C G C C
Y P I R E R Y N P I Q G Q R M V L K Y Y V T L H A L R R N I Y A S P P R S L E A
G T C A A A A A C T T A T A A T A C C A A T T T G G G T G A T A A T C T G G C G A T G T G C A T T G C G T C C A A A C C G G C C A C C C A G C G A T G G C C G C T A T T C C A T T C A G C A A T G
T M V F K N I O L K P S T I Q R H A D A I I R G F G A M G S P M A R R N W E A I
G A A C A C A C A C G C C T C C G C G A C A C G C G A G A C G G T G T C A T A T A C A A C A G T T C T A A A G T T T G C G G G C A A A T C G T A A T A A A T T G A C A G T T G T T T T T C A A C C A A G T C G A C T
S V C S A E A A A R S V T D D Y L L L E V L K R A Y O N I F Q C N N K R V L D V
C C C T C A T T A A A A C C T T C C G C C A A A T A C C C A A T G A C T T T T T C T T G T T A A A G C A A A G T T T A A A A G T T T T T C A C A C C C A C T T T A G G A G G A C A A A A C A G A C C G T T
G K M L V K G G Y I V G I Y K K K T I L L L K I F T K E C E L K I P P C F L A T
G A A A T A T A T G T G C A T T T T C G C C G A T T T A G C T A C C C T C A A C A C T A A C A C C C T G A A T C G G A T A A C A C A G A A T C C G A T C C A T A T A A C C T T A C C T G A C G C T T T T T G
S I I H A N K E G I K A I G E V S V G K F R I F V S D T O G Y I V K V E Y A K Q

VZV DNA sequence

1779

PSVVFV FHGGKHVLPSSAAPNLTRACNAARERF GFSRCQGP 128
CCATCTGCTTTGCTTTACGGAGGCAACAGTTTACCCAGCTCCGCGGCCCAATCTCACACGGCGGTAAACGGGCTGAGAACGGTTTGGGTTTTCACGCTGCCAAGGGCCT 51240

PV DGA VET TGA E I C T R L G L E P E N T I L Y L V V T A L F K E A V F M 168
CCTGTTGACGGTCTGTTGAGACGACCGGCTGAGATATGACCCGCTTGGATTAGAGCCAGAAATACAATATTATCTTGGTGGTACGGCATTGTTAAGGAAGCGGTATTTATG 51360

CNVFLHYGGGLDIVHINHGDVIRIPLFPVQLFMPDVNRLVP 208
TGCACGCTGTTTGCATTATGGAGGACTCGATTGTTATATTAACCATGGGATGTTATACGTATACCGTTATTTCCGGTACAACCTTTTCATGCCGATGTTAACCGTCTGGTACCC 51480

DPFNT HRS I G E G F V Y P T P F Y N T G L C H L I N D C V I A P M A V A 248
GACCATTCACACTATCAGGCTATCGGAGAGGGTTTGTATACCAACACCTTTTATAACACCGGGTTGTGCAATTAATACATGACTGTTTATGCTCCATGCGCTTGCC 51600

LRVRNVTAVARGAAHLAFDENHEGAVLPDITYTYFQSSS 288
TTGCGCTGAGAAATGTAACGCTGCGCGAGGAGCGGCCACCTTGCTTTTGATGAAAAACAGGAGGGGAGTACCCCCCTGACATTACGTACAGTATTTTCAGTCTCTTCA 51720

SGTTTARGARRNDVNSTSKPSPSGGFERRLASINAAADTAL 328
AGTGAACCATACCGCGGTGAGGCGCTGAAACGATGTCACTCCAGCTGTAGCCATCGGGGGGTTTGAAGACGGTTGGCGTCTATTATGCGCTGACACAGCCTTG 51840

HA E V I F N T G I Y E E T P T D I K E W P M F I G M E G T L P R L N A L G S Y 368
CAGCAGAGTATATTAACACTGGAATTACGAAGAACTCCAACAGATATCAAGAAATGGCAATGTTATAGGATGAGGGGACCTTGCCAGGCTAAACGCTGCGGCTCATAT 51960

TARVAGVIGAMVFS P N S A L Y L T E V E D S G M T E A K D G G P G P S 408
ACCGCTGCTGGCGGGGCTATTGCTGCGATGGTTTTCAGCCCAAAATCTGCGTTGTATCTAAGTGGAGGATAGCGGGATGACGAGCCAAAGGATGGGGGACCGGGTCCATCA 52080

FNRFYQFAGPHLAANPQTD RDGHVLS SQSTGSSNTEFSVD 448
TTTATCGATTATTACAGTTTTCGCGACCTCATTTAGCTGCAATCCCAACAGATCGAGATGGCCAGCTTCTATCCAGTCAGTCTACGGGTTTCATCAACACAGAGTTTAGCGTGGAT 52200

Y L A L I C G F G A P L L A R L L F Y L E R C D A G A F T G G H G D A L K Y V T 488
YATTGGCACTATTGCGATTGAGGACCCCTGTTGGCGGACTGCTTTTATCTAGAACGCTGTGACGCTGGTGGCTTACAGGGGGTACGGGGATGCGTTAAATATGTTAGC 52320

GTFDSEIPC SLCEKHTRPVCAHTTVHRLRQRMPRFGQATR 528
GGGACCTTGACTCGAAATTCATGATGTTTATGAAACACACGCGCGGTATGCGCTCACACACAGTACACCGACTTAGACAACGATGCCGCGATTGGACAAGCCACCGCT 52440

QPIGVFGT M N S Q Y S D C D P L G N Y A P Y L I L R K P G D Q T E A A K A 568
CAACCTATTGGGGTGTGGAACAAAGACGCAATATAGCGATCGCATCTCTAGGAACTATGCTCATATTAACTCTCGAAAACCGGGGATCAACGGAAGCAGCAAGGCA 52560

TWODTYRATLERLFI DLEQERLLDRGAPCSSEGLSSVIVD 608
ACCATCGAGCACTTATAGGCTACACAGACGCTGTTTATGATCTAGAACAGAGCGACTACTGGATCGCGGTGCCCCATGTTCTTCCGAGGACTATGCTCTGCTATTGCGAT 52680

HPTFRRI L D T L R A R I E Q T T T Q F M K V L V E T R D Y K I R E G L S E 648
CATCCAGCTTTCGTCGATATTAGACACACTGGTGGCGGTATAGAACAGACAAACAAATTTATGAAGGTGGTGGTTGAGACCGCGATTATAAGATCCGTGAAGGATTATCCGAA 52800

ATHSMALTFD P Y S G A F C P I T N F L V K R T H L A V V Q D L A L S Q C 688
GCCACCCATTCATGGCTTAAGCTTGTATCCACTACAGGAGCTTTTGTCCCATACCAATTTTATAGTTAAACGAACACACCTAGCCGTTGATACAGCTTAGCATTAAGCAATGT 52920

HCVFYGQQVEGRNFRNQFPVLR R R F V D L F N G G F I S T R S I 728
CATTTGTTATTTACGAGCAGAGTGTAGGGGGGAACTTTCGTAACCAATCCCACTGTTTTCGCGCGGCTTTTGTGACTGTTTAAATGGGGGTTTATATCAACACGCTCTATA 53040

YVTLSEGPVSAPNPTLGQDAPAGRTFDGGLARVSVEVIRD 768
ACCGTAACATTACTGAAGGCTGATTCGCCCCAAATCCGACATTGGGCAAGACGCGCCCGGGGCTACTTTGATGGGGATTAGCGCGGTAAAGCTGGAAGTTATTCGGGAT 53160

IRVKNRVVFSGNCTNLSEAAARARLVGLASAYQRQEKRVDM 808
ATACGAGTTAAAAATAGGTCGTTTTTCAGGTAACTGTACAAATCTCTGAGGCGCCCGGCAAGGCTTGTAGGCTTGAAGTGGTACCAACGCCAAGAAAAAGAGTGGATATG 53280

LHGALGFL L K Q F H G L L F P R G M P P N S K S P N P Q W F W T L L Q R N 848
TTACACGGGGCTAGGTTTTTGTCTAAACAGTTTACGGCTGTTATTTCTCGGGGTATGCCACCAACAGTAAATCCCCCAACCCGAGTGGTTTGGACCTGTTACACGCAAC 53400

QMPADKLT HEEIT T I A A V K R F T E E Y A A I N F I N L P P T C I G E 888
CAGATGCCGAGATAAATACACGAGAGATTACCACTATTCAGCTGTTTAAACGGTTTACCGAGGAATGACGAAATAAATTTTATTAATACCCCCAACCTGCATAGGAGAA 53520

LAQFYMANLILKYCDHSQYLINTLTSIITGARRPRDPSSV 928
TTAGCCGAGTTTATATGCGAAATCTATTCTAAATACTGCGATCATCACAGTACCTTATAAATACCTTAACCTCTATAATTACGGGTGCGAGGCGCCGCTGACCCATCATCCGTT 53640

LHWIRKDVTSAAD I E T Q A K A L L E K T E N L P E L N T T A F T S T H 968
TTGCATTGGATCGTAAAGATGTCAGCTCCGCGGAGATAGAAACCAAGCAAGAGGCGCTTCTTGA AAAACGGAACCTACC GGAATTATGGACTACGGCTTTTACTTCACTCAT 53760

LYRAAMNQRPMVVLG I S I S K Y H G A A G N N R V F Q A G N W S G L N 1008
TTAGTCCGCGGCGCATGAATCAACGCTCCATGGTGTTTAGGAATAAGCATAGTAAATATCACGAGCGGCGAGGAACACCCGCTTTTCAGGAGGGAATTGGAGCGTTTAAAC 53880

G G K N V C P L F T F D R T R R F I I A C P R G G F I C P Y T G P S S G N R E T 1048
GGGGTAAAAATGATGCGGCTATTACATTGATGCGACTCGCGTTTATAATAGCATGTCTAGAGGAGGTTTATCTGCCCGTAACAGGTCCTGTCGGGAAATCGAGAACC 54000

TLSDOVRGIIVS G G A M V Q L A I Y A T V V R A V G A R A Q H M A F D D 1088
ACCCTATCCGACCAAGTTCGCGGTATAATTGTCAGTGGCGGGGCTATGGTTCAATTAGCCATATACGCCACGGTTGTCGTCAGTGGGCTCGAGCACAACATATGGCATTGACGAC 54120

	WLSLTDDDEF LARDLEELHDOI IOTLETPNTVEGALEAYKI	1128
	TGGTTAAGTCTTACAGACGATGAGTTTATGCCAGAGACTTGGAGGATTACAGACCAGATTATCCAACCCCTGGAACGCCCTGGACCGTAGAAGGCCCTCTAGAAGCAGTAAAGATT	54240
	LDEKTTYAGDGETPTNLA FNFDSC E PSHD TTS NVL NIS GSN	1168
	CTAGATGAAAACACGACGCGGAGATGGGGAACCCCAACCTAGCATTATATTTGATCTGTGGAACCAAGCCATGACACCACTCAACGATTAAACATTTCAAGGTCACAC	54360
	ISGSTVPG LKRPPEDELF DLSG IPIKHGNITMEMI -	1204
	ATTCAGGGTCAACTGTCCCTGGTCTTAAACGACCCCGAAGATGACGAACCTTTGATCTTAGTGGTATCCCAATAAACATTGGAACATTACAATGGAATGATTAACTCCCTCT	54480
	TTATCCAATTAAGCCACACGCGGGTGAGTGACGTAATAACAAGTCAATATTACATATCTGTGTGTTTCTTTTTTGTGTGTGTGCTTACCCATGACCTGTAATATAGTG	54600
30	MELDINRTLLVLLGQVYTYIFQV	23
	GTCTCAACCATTCAGCTYACAGTCCAGTGAGACGTAACAGCCGATACATGGAATTGGATATTAATGAACATTGTTGGTTCTACTG66TCAAGTTATACGTACATCTTTCAGGTTG	54720
	ELLRRCDPRVACRFLYRLAANCLTVRYLLKLFLRGFNTQL	63
	AACGTCTACGTGATGATCCAAGGGTGGGTGTCCTTTTATATCGGTAGCGCTAAGTGTGACAGTTCGTTATTTATTAAGCTGTTCTCCGGGATTAATACCCAGCTAA	54840
	KFGNTPTV CALH WALCYVKG EGERLFELLQHFKTRFVYGE	103
	AATTTGAAACACTCCACGGTTTGTGCACTGCATTGGCATTATGTTATGTAAGGGAAGGTGAGCGTTTGTGTGAGTTGCTACAACATTTAAACCGGTTTGTATTGTTGAGA	54960
	TKDSNCKIKDYFVS AFNLKTCQYHHELSLT TYGGVVSSEIQ	143
	CTAAGACTCAACTGTATCAAGATTACTTGTCTACGCTTAACTAAAACCTGCCAATATCACCAGAGTGTGTTAAACATACCGAGGTACGTACGAGTGAATTCAGT	55080
	FLHD IENFLKQLNYCYIITSSREALNTLET VTRFMTDTIG	183
	TTTTACACGATGAGATTTTAAACAGCTTAATGCTATATACGCTTCTCGTGAGCGCTAAACACATGGAACCGTGACCGGTTATGACAGATACTATAGGAA	55200
	SGLIPPVELFDPAHPCAICFEELCITANQGETLHRRLLGC	223
	GCGGTCTAATACCCCGTGGAGTTGTTGATCCGGCGATCCATGTGCTATATGTTTGAAGAATTATGTAACAGCTAACCAAGGTGAGACCTTACATCGTAGATTATAGGATGTA	55320
	ICDHVTKQVRVNVDDIIRCLPYIPDVPDIKROS AVEAL	263
	TCGCGATCAGCTTAAAGCAAGTTCGGGTAAACGGATGTTGACGATATTATCGGTGTTACCATATCCCTGATGACCGGATCAACAGCTCAATCCGCGTGAAGCGTTAC	55440
	RTLQTKTVVNP MGAKNDTFDQTYEIASTM LDSYNVFKPAP	303
	GAACACTTCAACCAAGCGTAGTCAATCCATGGGAGCAAGAAGATACGTTTGACCAACATACGAAATGCGAGCACCATGTTGATTCTTAATGTTTAAACCTGCCCTC	55560
	RCMYAISELKFWLTSNSTEGPQRTLDV FVDNL DVLNEHEK	343
	GGTGTATGTACGCCATCAGCGAGCTAAATCTGGTAAACGCTAATTCACGAGGACCCCAAGCTTATAGAGCTGTTGTTGATAATTTGGATGATTAACAGCAATGAAAAAC	55680
	HAELTAVTVELALF GKTPIHFDRAFSEELGSLDAIDSILV	383
	ACGAGAACTTACAGCGTAACGGTTGAGTTGGGCTATTGGAAAACTCCATACACTTGTAGGGCGTTTCTGAAGAACTCGGATCTCGGATGCAATGATAGTATTGTTGTTG	55800
	GWRSSSPDSQIEALIKACYAHHLSSPLMRHISNPSHDNEA	423
	GCAATGCTCATCTCACCAGAGCTAGATAGAGCAATTAATAAGCTGTTATGCCATCATCTATCGTCCCTCTCATGCTCACATTTCAACCGAGTCATGATAACGAAGCG	55920
	ALRQLLERVYG CEDDLTKEASDSATASECDLNDSSITFAV	463
	CCTTACGCCAATTTAGAAAGAGTGGGTGAGGATGATTAAACCAAGAGGCGAGTGACAGCGCTACAGCATCGGAATGATGATGACAGATAGTAGCATACCTTTGCTGTTT	56040
	HGMENLLSKAKIDAAERKRVYLEHLSKRSLSLGR CIREQ	503
	ATGGATGGGAAACCTGTTATCCAAGCAAAATGACGCTGCGGAAAGAAACGAGTATATCTGAACATCTGCTAAGCGCTCTCAACCGCTCGGATGATGATCCGCAACAGC	56160
	RQELEKTLRVNVYGEALLQTFVSMQNGFGARNVFLAKVSO	543
	GCCAAGAGCTAGAAAAACACTCAGGGTAAACGTTTATGGAGAGGCTTATGACAGACTTGTTCGATGCAAAATGGGTTGGGGCACGAAACGTTTATAGCTAAGGTTCCCGAG	56280
	AGCIIIDNRIQEAAFD AHRFIRNTLV RHTVDAAML PALTHK	583
	CAGGGTGATTATCGACAATCGATTACAGGAAGCGGCTTGTGACATAGATTATAAGGAATACCTTAGTTCGACATACAGTAGATGCGGCTATGTTACCTGCCTTACACATAAT	56400
	FFELYNGPLFNHDEH RFAQPPNTALFFT VENVGLFPHLKE	623
	TTTTGAGTTGGTCAACGGCCATGTTTAACTACGATGAACACCGTTTGCACACCCCTAACACCGCTTATTTTTACCCTGGAACGTTGCTTATTTCCGACTTAAAGAGG	56520
	ELAKFMGGVVGSNWLLSPFRGFYCFSGVEGVTF AORLAWK	663
	AATTTGCAAGTTTATGGCGGTGTGTTGGTTCCACTGGCTTCTAGTCCATTATGAGGCTTTTATGCTTTCTGGGTAGAAAGCGTTACTTTTGCACAGAGCTTGGCTGGAAT	56640
	YIRELVFATTLFTSVFHCGEVRLCRYDRLGKDP R GCTSQP	703
	ATATTAGGAGCTGTGTTTGCACACACTATTCACCTGTTTTCATTGTGGGAGGTGCGGTATGTGCGGTGACCGCTAGGTAAAGATCCACGCGGTGACGCTCAACCTA	56760
	KGIGSSSHG PLDGIYLT YEETCPLVAIIQSGETGIDQNTVV	743
	AAGGTATAGGCGTTCCACGGACCTTAGACGGCATTATTAACGTACGAAGAAACATGCCCCCTGTGGCTATTATCAAGTGGAAGAACGAGGATGACCAAGATACCGCTGATA	56880
	IYDS DVFSLLYTLMQRLAPDSTDP AFS -	770
	TCACGATTCAGAGCTTTTCTCTATACACCTAATGACGCGCTGGCTCCGGATTCAACGGACCGGCTTTTCATACCTCCGTTACGGGGGTGTTGTTATGCTTTTATGAT	57000
31	MFVTA VSVSPSSSFYESLQVEPTQSE DITR SAHLG DGD	38
	ATTTCTATGTTGTTACGGCGGTGTGTCGGTCTCCAAGCTCGTTTATGAGGTTTACAAGTAGAGCCACACAATCAGAAGATATAACCCGCTGCTCATCTGCGGATGGTGA	57120

VZV DNA sequence

1781

E I R E A I H K S Q D A E T K P T F V V C P P P T G S T I V R L E P T R T C P D 78
TGAAATCAGAGAAGCTATACACAGTCCCAGGACGCCGAAACAAACCCACGTTTACGTCTGCCACCGCCACAGGCTCCACATCGTACGATTAGAACCACTCGGACATGTCGGGA 57240

Y H L G K N F T E G I A V V Y K E M I A A Y K F K A T V V Y K D V I V S T A W A 118
TTATCACCCTTGGTAAACCTTACAGAGGGTATTGCTGTTTATAAGAAAAACATTGACGCTACAGTTTAAGGCGACGGTATATTACAAAGATGTTATCGTTAGCACGGCGTGGGC 57360

G S S Y T Q I T N R Y A D R V P I P V S E I T D T I D K F G K C S S K A T Y V R 158
CGGAAGTCTTATACGCAATYACTAATAGATATGCGGATAGGGTACCAATTCCTGTTTACAGATACGCGACACCATTTGATAAGTTTGGCAAGTGTCTTCTAAAGCAACGTACGTACG 57480

N N H K V E A F N E D K N P Q D M P L I A S K Y N S V G S K A W H T T N D T Y M 198
AAATAACCAAGTTGAAGCTTTAATGAGGATAAAATCCACAGGATATGCCCTAATCGCATCAAAATATAATTCTGTGGGATCCAAAGCATGGCATATACCAATGACACGTACAT 57600

V A G T P G T Y R T G T S V N C I I E E V E A R S I F P Y D S F G L S T G D I I 238
GGTTGCCGGAACCCCGGAACATATAGGACGGGACGTGGTGAATTGCATCATTGAGGAAGTTGAAGCGAGATCAATATTCCTTATGATAGTTTGGACTTCCACGGGAGATATAAT 57720

Y M S P F F G L R D G A Y R E H S N Y A M D R F H Q F E G Y R Q R D L D T R A L 278
ATACATGTCCTGTTTGGCTACGGGATGGTGCATACAGAGAACATTCGAATTATGCAATGGATCGTTTCCACAGTTTGAAGGTTATAGACAAAGGGATCTTGACACTAGAGCATT 57840

L E P A A R N F L V T P H L T V G W N W K P K R T E V C S L V K M R E V E D V V 318
ACTGGAATCTGACGCGGGAACCTTTAGTCAGGCTCATTAAACGGTGGTGGGATGGAAGCCAAACGGAAGTTTGTTCGCTTGTCAAGTGCGGTGAGGTTGAAGACGTAGT 57960

R D E V A H N F R F T M K T L S T T F I S E T N E F N L N Q I H L S Q C V K E E 358
TCGGATGAGTATGACACAAATTCGCTTACAAAGAAACACTTCTACACGTTTATAAGTGAACAAACGAGTTTAACTTAACCAATCCATCTCAGTCAATGTGTAAGGAGGA 58080

A R A I I N R I Y T T R Y N S S H V R T G D I Q T Y L A R G G F V V V F Q P L L 398
AGCCCGGCTATTATTAACGGATCTATCAACACAGATACAACCTATCTATGTTAGAACCGGGGATATCCAGACCTACCTTGCAGAGGGGGTTGTGTGTGTTTCAACCCCTGCT 58200

S N S L A R L Y L Q E L V R E N T N H S P Q K H P T R N T R S R R S V P V E L R 438
GAGCAATCCCTGCCCGTCTCTATCCAGAAATGGTCCGTGAAACACTAATCATTACCAACAAACACCCGACTCGAAATACAGATCCGACGAAAGCTGCAAGTTGAGTTGGC 58320

A N R T I T T T S S V E F A M L Q F T Y D H I Q E H V N E M L A R I S S S W C Q 478
TGCCATAGACAAATAACACCCATCATCGGTGAATTTGCTATGCTCAGTTTACATATGACCACATTCAAGAGCATGTTAATGAATGTGGCAGGTATCTCTCGTGTGGTGGCA 58440

L Q N R E R A L W S G L F P I N P S A L A S T I L D Q R V K A R I L G D V I S V 518
GCTACAAATCGGACGCGCCCTTTGGAGCGGACTATTCCAATTAACCAAGTGCTTATGCGAGCACCATTGTTGGATCAACGTGTTAAAGCTGATTCCTCGGCGAGTTATCTCCGT 58560

S N C P E L G S D T R I I L Q N S M R V S G S T T R C Y S R P L I S I V S L N G 558
TTCTAATGTCCAGAGCTGGGATCAGATACAGCATATATCTCAAACTCTATGAGGGTATCTGGTAGTACTACGCGTTGTTATAGCGGCTCTTAATTTCAATAGTTAGTTAAATGG 58680

S G T V E G Q L G T D N E L I M S R D L L E P C V A N H K R Y F L F G H H V V Y 598
GTCCGGGACGGTGGAGGGCAGCTTGAACAGATAACGAGTAAATATGTCCAGAGATCTGTAGAACCATGCGTGGTAAATCAAGCGATATTTCTATTGGGATCACTACGTATA 58800

Y E D Y R V V R E I A V H D V G M I S T Y V D L N L T L L K D R E F M P L Q V Y 638
TTATGAGGATATCGTTACGTCGGTGAATCGAGTCCATGATGTGGGAATGATTAGCACTACGTAGATTAACTTAACACTTCTTAAGATAGAGAGTTATGCCGTCGAAGTATA 58920

T R D E L R D T G L L D Y S E I Q R R N Q M H S L R F Y D I D K V V Q Y D S G T 678
TACAAGAGACAGCTGGGATACAGGATTAAGTACAGTACAGTGAATTCACGCGGAAATCAAAATGATTGCTGCTGTTTATGACATAGACAAGTTGTCAATATGATAGCGGAAC 59040

A I M Q G M A Q F F Q G L G T A G Q A V G H V V L G A T G A L L S T V H G F T T 718
GGCATTATGACGGGATGGCTAGTTTTCAGGGACTTGGGACCGGGCCAGGCGTTGGACATGTGGTTCTTGGGGCCAGGGAGCGCTTTCCACCGTACACGATTATACCAAC 59160

F L S N P F G A L A V G L L V L A G L V A A F F A Y R Y V L K L K T S P M K A L 758
GTTTTATCTACCCATTGGGGCATTTGGCGTGGGATATTGGTTTGGCGGACTGGTAGCGGCTTTTGGCTACCGTACGTGTTAACTTAAACAAAGCCGATGAAGGATT 59280

Y P L T T K G L K Q L P E G M D P F A E K P N A T D T P I E E I G D S Q N T E P 798
ATATCCACTACAAACAGGGGTTAAACAGTTACCGGAAGGAATGGATCCCTTGGCGAGAAACCAACGCTACTGATACCCCAATAGAAAGAAATGGCGACTACAAAACACTGAACC 59400

S Y N S G F D P D K F R E A Q E M I K Y M T L V S A A E R Q E S K A R K K N K T 838
GTCCGTAATAGCGGGTTGATCCGATAAATTCGAGAAGCCAGGAATGATTAATATATGACGTTAGTATGCGGCTGAGCGCAAGAACTAAAGCCGCAAAAAAATAAGAC 59520

S A L L T S R L T G L A L R N R R G Y S R V R T E N V T G V - 868
TAGCGCCCTTTTAACCTTACGCTTACCGGCTTGTTCAGAAATCGCCGAGGATATCCCGTGTGCGACCGAGAAATGAACGGGGGTGAAATAGCCAGGGGGTTGTTTTAATTA 59640

TTAATAAATATGTTATACGTTACTGATGTCCTCATTACGATACAGGGGGATTATATACCGGATAATATACAAACGCGTTTGTACCTCTACCGCACCGGATATCTAACGGGG 59760

32 M E S S N I N A L Q Q P S S I A H H P S K Q C A S S L N E T V K D S P P A I 38
TTATATGGAATCGCTAACATTAACCGGCTACAAACACCGTCTATCGACATATCCGTCACAAACAGTGGCTTCAAGTCTCAATGAACAGTAAGAGATTCTCCCCCGCGATT 59880

Y E D R L E H T P V Q L P R D G T P R D V C S V G Q L T C R A C A T K P F R L N 78
ATGAAGATAGGTAGAACACACGCGGTAATACCCGCGACGGTACACCCGAGAGCTATGTTCTGTGGGACAGTAACTGTGAGCATGTGAACGAAACCTTTTCGCTTAACC 60000

R D S Q Y D Y L N T C P G G R H I S L A L E I I T G R W V C I P R V F P D T P E 118
GCACAGCAATACGACTACTTAACACATGTCCAGGGGGCGCTATATTCACCTGACCTGGAGATTATAACGGGTCGATGGGTTGTCATCCCGGCTGTGTTCCGGATACCCAGAGG 60120

	E K M A P Y I I P O R E Q P S S G D E D S D T D -	143
	AAAAATGGATGGCCATATATATCCAGACCGAGAACCAACCATCATCAGGGGATGAAGATCTCGACCCGATTAATTAACCTTAATAAACCTTACCACCCATAAAAAACCCCTTCT	60240
33	GTGTGTTAACACGACACCGCTTAACAAAAAACCACCAACGCCCTTTATGAATGAATATTTTGTGGTGAACACCGCCACCATCATCTGATTGCAACATATCG	60360
	- C R G . V M M Q N A F M D	594
	CGTCGTCTCCGTCGACCCCTGTATTAAGGGGCTTGGAACTCGCTCCACTGCATTACATCTTGCCAACGTATCTGTATGTGGGGTCTGTGTATTTGGGATGAGCATAGA	60480
	A D D A T S G Q I L P A K S S A E V A N V D Q G V T D T H P T S T T N Q S S C L	554
	CCCGAAACGCTTGAAGCTGTTTAAATAAATCGATATTCGAGGATCCCGTCTCCCTCTGGTATATTTGTATGGTGACAAAGGCATTGTGCCATTTGTGATTTAGCTCTGTA	60600
	G S Y S Q L Q K L L I S I R P D R T G E P I N T H S L P M Q T G N Q S K L E T	514
	ACCTCTGTGCACTTTGCCACACCCAGCAAGCTCTCTGCTGACCATAGAACTCTGTCTCTCTGCAATATGATGGAGAACTCGAGCTCTCCGATGCGTATATACGTT	60720
	V E Q Q L K A V V G A L E E H Q G N S V R H R R Q M Y S P S V R R R R H T I Y T	474
	GTTTACCAGGAAATATATTTGAGGAAACTCTCCGCTCATTTGAGACTCCCACTATAAAGAATCAATTCCTTTGATCCATGCTTGTAAATCCGCTTTCTGAGCAGCGG	60840
	P E G P F Y I N S P F E G D M Q S E G S Y F S D L E R Q D M S K F D R K G P R R	434
	ACATCGGTTTGTCTGGAATTTACACACGGGCTCGCAAGTCAATACCCGCTCGGCGGCAATGCGTTCAATGCGGACATTGATTTCCAAACGATTGGGTGGTGATATCCC	60960
	V D T K D P F N V C P T Q L D I G R E A A L A N M F A S M Q M E L R N P P P Y G	394
	GGAAACCGTACCGTCCCCGAAGTGTCCGAGGGAACCAATAACCCCTGTATTAAGTGGGAAGCAGCGGGTGGAGATCCATATGCGGACGATATCTGCGTATTGGA	61080
	P F G Y P G G F H G P P C G Y G G T N P P F A P P H P S G Y P G V I Y Q G N N P	354
	GTCGAATGATACCTGCGGATTTTGTCTGCGGTTAACAGCTGTGAATAATACGCGGTAGGTACGATCAAAATCCCTCGGTTGGAAGCGGACGCGGCTGTGGTGAGATA	61200
	A G I S V Q P N K T Q G T L L Q S Y Y A T P I L V F E G G T P V G S P P Q P S I	314
	TTACTAGCTTACCTGCTACAGAACCATATCGTGTCTCTACCAACTGCGTAACCTTTAAATGCGGAACAGCTTTTCAAACTTTCATTTGATTTCCCAACCCACCGGAGA	61320
	N S A N G A V S A M D S D N R C L Q T V K L H P V T K E C D E N S E G V G L A L	274
	TGCTATATGGGCGCGGGTGAATGTGGCTTTTAAACCCCGGTTGGTAAATTTAGACTCCACCCATTAACGTTGGTATCCGAGCAAGTCATATCCGCTGATGCTGAAGA	61440
	D Y I P G G P H F T A N I V G A N P L K S E V G N V N T I R A L G Y G T S A Q L	234
	TAACTGACCCCAATTCGCGTCTCGCTACGTTTGAACACAGTCCCATCTATCTTAAAGCATATTTGTCACGGCTGGATATAACACCTTGGCGAGTTATCTTCGTA	61560
	Y V H G M . J G A E R R R K A V V D W R D R L L M N N V A T S L L V K A L K D E S	194
	ACCTTCCATCTTATTTAAACCCCGTGTGTTTAAACAGCGACAATAACCGCGCTTACTTTCCATCGATAAAACCCGGAATGGTCAATTAAGATTCGCGGATACAGTCATAATG	61680
	V K N V K N L G A Y D K V L S L L R A K S E M S L V R F P E I S S E P T C D Y N	154
	ACCCTGTTCACCGCTTCTCAACAACACATAACGCAACATGGGTAAAAAATACCGTCTGATCTCATTGCGGACAACTCGTTTGAAGACAGGATACGAGGGTAAGTAATTT	61800
	V V T G V R R G V V C L A V H T F F N G D P I E N P S L R K S S L S V S P L Y N	114
	GTGACCAAGTATAACGACGTTCTAGCGGAGATAATACAGATCTCTATTTCCAAAAAATCGAATGGGCGCTTCAACAGCAGCCGATGATGTAGGCGATCAACGATACCCAA	61920
	T V L Y L A R E L P S L V S D R N G F F N S H A A E F L V A H L Q P C R V I G L	74
	AAAAAGGTCCCGTATGCTCAATGATGCGATTACTTACCCAGCAGCAGCTTTTGCATGATGATGTTTATGTTTACTAGTAGGCGGCAAGCGGACCGCAATCTCT	62040
	F F P G R I D E I I A I V E G V V C D K R H D I N I P I K S T P P L A S R V I E	34
	GGGGTAATTTAAATCCCTTCGCTTGAATATAAGGTAATACCCAGCAGTATAACGCTTACAGTCTCTCTGTCAGCTTCAGCAGCTTATAAACACCCACGAGCCGGA	62160
	P T I N L E G E D K S Y L A L Y G A V Y L A E C N E E D A E A A M	1
34	TAGTGAATCTACCGGTGTGGAGGCAACTGAGGAATGACACCCAAACAGACAAAATAGAAGATCATAGTCACTGTTAACGTTGAACGAGGCGGACCTTCTTCCAATGCCG	62280
	- P T S A F Q P I V G L C V F Y L L D Y D S N V N F . Q A L R R S E E L A A	544
	CCCTTACACCGGTTGGTGATTAACATTCAGTCCCGTTCATATTGCAACATAACACTGTATGATTGATACCGGCGGCTATGGGTAGGATGATCAATTTTGTGCGGCTGT	62400
	R V R P Q H M L M G L G R E Y Q L M V S D H I S V V A A I P L S T V N Q R R H E	504
	CTAATTCGAATTAAGCTTATGAGCGATCTTGTACTGTCCAGAGAATAATCTATTACGGTCTTCTTAACTTCCAGACTAAGCTGGGTATGCGGCTTAACAAAGAGCAA	62520
	L E L A I L S I L R D Q Y Q G S S I D I V T R G L S G R S L Q T H A D L C L A V	464
	CTAATCCAGGAACATTTAGTCAGCTGTGGTCCGATTTAAGTATACAGTGGTCTATATCGTTCACATAAAATGAAGTATTATTACCCTTTTAACTTCCCATCAAAC	62640
	L G P F M E T L E T T R N L T Y L P A I Y R E C L F Q F N N N G S K F K G D F G	424
	CCGTCGCTCCGCGCAAGATTACATTGTTGGTAGGGGTTCTGTTCTGACACAATAAACCCAGTGAAGAAATATTTTATGTTATCTCCGATACGTTCCGTTCCATAAAGC	62760
	T A G R L I V N N T P T G T A E S V I L G L Q F N K L K D G Y V N G N M L L R	384
	GCCTTAATAATAACCGGTAATCGTGTCAATTTGTAACCTTAATAGAGTTGGTCTTCCATAAGAAACAGTGTGGGCGGCTTAAATACGCGCGGCGCTGTTGAATCTGT	62880
	R L L L L A T I T D I T L R L L T Q D E M L F V N Q A R E L Y A A A A Q Q I K D	344
	CCACATATGCGGTATGATGCGATCAATATGTCATTAACCCAGGATTAACCTGTCAGGTGACGCGGAGGACCTGCAACCGTATAAGCGCATCAACAGAAATGTGACGTTAAG	63000
	V Y A T H N R D I I D N V G P N F Q G P A P P L Y Q L R I L A D M V S H S T L A	304
	CGCTTGTATACCGCCCCCAGCAGTGAAGTGGTGGGTAGAGCATATAGCAAAATGATAACTGTTTATTTTCTGTTGTGTCATATAATTCACAAATGTCTCAGTAT	63120
	G Q O Y R G G R A H F Q D R P L R D Y C F Q Y S N K N E H T T M Y N V F T E T Y	264

1783

143	60240	ATTCCGGTAGTGTCTATAAGGTTCCCGAAGCAAACTTGGAGGACTATTAGTAGTCTATACATTAATAAACAATATCCGACACATCGAACCGGAGTACGCTC	63240
6	60360	EPLNHEILNGFSSVQPEHVSNSTRYMLYLCLLVACEFASVAR	224
594	60360	TATCTCCAACATACATTCTCCCGGGGAGTGTAGACATGTTACGGTGTGTTTCATTAACAGTACGGGAATGCGCCGCTTTTCAATCAACTCCGCGTGCAGTACGGGCTATCTAACA	63360
60480	554	DGVYMRGASQLCTVTTNMFTRSGDKCDVGRAAVPRDLV	184
60600	514	CAAGCGTTCCTGCAGAGTACGATACCATGCCCCGAAAAAATCCCTGGAGAGTTATTGCCCTTGGCTTCCCAAGTACACAGGGTGTATTAATCCACTTGAAGTTGTATCTGCTAT	63480
60720	474	LREQLTRYWPGFVIGPSNNGRARGLYVLTIFDVQFNTDYO	144
60840	434	GCACGGTGCATCTTTTGGCAATCTGTACCTCGGGGTGTATAGACTCATTGGGTATTATTTCTGTACGTGTACATCTCCAGATTGTCATCTGCTTCTTCCGCTCGGACGACGGC	63600
60960	394	LPADNKAIQVEPHISENRIIETRTCEESQADAEAEAAAT	104
61080	354	TCTCCAGGAATCCAAACCTTGGCCATGCGGTAGTTGTTCTCGAGGGCTTTAAACAGCATCTATTTCCGTGGTAACTGATCTTCCCGCGAAGGTGTCTAATGCGGCAA	63720
61200	314	ELSDLVKAMRTLQEEELPKLLRRDIETPLTITEGRLNDLAAV	64
61320	274	CGCGCGCGCATTTTAAAGTTAAAGTATTTTTTCCAAATCGGGATCATACGCCCTTAACTCAACAGCGGGAGCGCTCAGTAGTGTATGGGAAGTTGGGGGTCTAAGAGTTCT	63840
61440	234	AAANKLTLTNKELDPNMWRGRLEFAPATWYHIFNPAIFNK	24
61560	194	TAGTGGTAGACAAAATATCCCATTTTATCGGAAGCAGATAGATCGAACCCATATCTCGCGTGTAGTGTGTGACGAAACAAAGTCACTGGCGTGAATATAACCCGGTACTGC	63960
61680	154	TTSLFLIGCKNPFSSISGFGYRATM	1
61800	114	TTTAAAGCTGTTTCTTACCATGGGAAACATCCCGGTTATCTTGTAAATTCACCAACAGCACCTAAAGAAGGCTTCAAGGGTAAATCCACCCCAAGCTGCATTTCTT	64080
61920	74	-GHSFMGTISQLIGGCAGLSPRRLPLDVGCAANE	225
62040	34	CAAACTTTGTAAAGCGGACGATGGCATGATTCGACGCTTTTCGCAAGAGAACATACGTGAATTTCTTTTGTACAGAGCTTCTGCTCTCAAGGACCTATATCGGGGGGTAT	64200
62160	1	FKTLASRHC SKARKKALSCVHIKKKCLRRRERVSRIPPTY	185
62280	544	ATTCGGTACATCTCCAAATCGGACGTAGCATACAAGGTTTCCATGAATCACCTTTGGGGTAAACGAGTACCTGTATCAAGGTTCAAGCCCGTGAATACAAACACAAGGAGGG	64320
62400	504	EAVNELHLSALMVLNGHIIVKPLRLTVQLNLGRQSVFVLLP	145
62520	464	CGGTCAACATATTTTCATCAGATCCGCTGGGTGGTTTCTTTTAAAGGCCATGGTATCCCTCAGTGGGCGATACCTTCGAAACTGGTGTACTAGTAGGGGTATGTATATAG	64440
62640	424	TYMIEDSGTPTTEKILAMTDRQLQRMGECFQHYKTPTHINA	105
62760	384	CGCTAAACCGCAAGATTTTAACTCCATATAAAACAAACGGTCTTTCCGGACCACTGGATTCCGTTGTATATAACAAACACATCGGGGCGTGGGCTCCCAATTTACTTCAACG	64560
62880	344	SFRCSKLEVIFCVTKGAGSSETQIICVCDPRRRGLNVEFS	65
63000	304	...	
63120	264	ACATTGATATCGGTACAGCCCTTTGAATCCACGTGGGTAACGGGACAGGAGTTTCCAGCCTCGGGTGAACGGTCCGCGAAACCTCGACGTACTATCAATATCCTTTTGA	64680
		SIRVARQVDVHSLPSLLKALRPNFADAFGRVNDIDKKL	25
		AccI	
		GTACATGTAACAGAGTGTGGCAAGGTTGCCAACGAAACACTTGGCCGAATTGCAGTACGGGACATATTGAAGTTCCGCCCAGAGATAACCTAAGACGCGTTTGTCTACA	64800
		VDYFRTHCRQGLRFCKARIRSAM	1
		MSTDKTDVKMGVLRIRIYLDGAYGIGKTTAAEEFLHHFAI	38
		ATAACATGTCAACGGATAAACCGATGATAAATGGGCGTTTGGCATTTATTTGGACGGGGGTATGGAATTTGGAATAACACCGCCGCGGAAGATTTTACACCATTTGCAATA	64920
		TPMIRILLIGEPLSYWRNLAGEDAICGIYGTQTRRLNGDVS	78
		ACACCAACCGGATCTTACTCATTTGGGGAGCCCTGTGCTATTGGCGTAACCTTGCAGGGGAGGACGCAATTTGGGAATTTACGGAACAAACATCGCGCTCTAATGGAGAGCTTTCG	65040
		PEDAQRRLTAHFQSLFCSPHAIWHAKISALMDTSTSDLVQV	118
		CCTGAAGAGCCACACGCTTACGGCTATTTTCAGAGCTGTCTGTGTTCCGCATGCAATATGATGATGGAATAATCGGCAATGATGGACACAAGTACATCGGATCTGTCAGTAAGTA	65160
		NKEPYKIMLSDRHPIASTICFPLSRYLVLGDMSPAALPGLL	158
		AATAAGGAGCGTATAAAATGATGATTCGCGACACCCCAATCGCTCAACTATATGTTTCCCTGTGCAGATACTAGTGGGAGATATGTCGCGACGCGCTTCTCGGGTATTG	65280
		FTLPAEPPGTNLVVCVTVSLPSHLSRVSKRARPGETVNLPPF	198
		TTTACGTTCCCGTGAACCCCGGGGCAACTTGGTAGTTGTACEGTTTCACTCCCGCATTTTATCCAGAGTAGCAACAGGGCCAGACCGGGAGAAACGGTAACTGCGCGTTT	65400
		XbaI	
		VMVLRNVYIMLINTIIFLKTNNWHAGWNTLSFCNDVFKQK	238
		GTTATGGTCTGAGAAATGTATATATAGCTTATTAATAACAATTATATTTTAAACATAACACTGGCAGCGGGCTGGAAACACATGTCATTTTGAATGATGATTTAAACAGAAA	65520
		LQKSECIKLRLEVPGIEDTLFAVLKLPPELCEGEGFNILPLWA	278
		TTACAAAATCCGAGTGATAAACTACGGGAAGTACCTGGGATGAAGACACGTTATTCGCCGTGCTTAACCTCCGAGCTTTCGGAGAGTTTGGAAATATTCGCGTTATGGGCA	65640
		NGMETLSNCSRSMSPPFVLSLEQTTPQHAQAQELKTLPLPQMT	318
		TGGGAAATGGAGCCCTTTCAACTGCTCACGAAGATGCTTCCGTTGATATGCTGTAGACAGACACCCAGCATGCGGCAAGAAGCAATAAATCTGCTACCCAGATGACCCCG	65760
		GCANACATGCTCCGGTGCATGGAATATATTGAAGAGCTTGTAAATGCCGTGAGGACAACATCTCTAAATATACCTAGTATTTACGTATGTACCAAGTAAAGATGATACACATTG	65880
		
		3' end of dPyk mRNA	
		TCATACGCGGTACGTTGTTTTCTTTTATATATGCGTCATTTATACCACATCTTTAATCCCGCTTTATCTCCCTAAACAGGAGTGGTAAATATAAAGCGGCAAGCTGTTG	66000

37	6TGGGTGAGGAGGGTAAAGGACGCTGTGTGATAACGTTGGGTGATATTGTAGCGCAAGTAACAGCGACTATGTTTGGCGTAGTTTATAGCGGTGGTAATCTTCTCTTTTGACCAC	16 66120
	A N K S Y V T P T P A T R S I G H M S A L L R E Y S D R N M S L K L E A F Y P T GGCTAATAAATCTTACGTACACCAACCCCTGCGACTCGCTCTATCGGACATATGCTGCTCTCTACGAGAATATCCGACCGTAATATGCTCTGAAATTAGAAGCCCTTTATCTAC	56 66240
	G F D E E L I K S L H W G N D R K H V F L V I V K V N P T T H E G D V G L V I F TGGTTTCGATGAAGAACTCATTAAATCACTTCACTGGGAAATGATAGAAAACAGCTTTCTTGGTTATTGTTAAGGTTAACCTTACAACACAGGAAGGAGCTCGGGCTGGTTATATT	96 66360
	P K Y L L S P Y H F K A E H R A P F P A G R F G F L S H P V T P D V S F F D S S TCCAAAATACTTGTATCGCCATACATTTCAAAGCAGAACATCGAGACCGTTCTCTGCTGGACGTTTGGATTCTTAGTCAACCTGTGACACCGACGAGCTCTTGTACAGCTTC	136 66480
	F A P Y L T T Q H L V A F T T F P P N P L V W H L E R A E T A A T A E R P F G V GTTTGGCCGATTTAATACGACATCTTGTGCTGTTTACTACGTTCCCAACACCCCTTGTATGGCATTGGAAAGAGCTGAGACCGAGCACTGCAAGAGCCGCTTGGGGT	176 66600
	S L L P A R P T V P K N T I L E H K A H F A T M D A L A R H T F F S A E A I I T AAGTCTTTACCCGCTGCCCAACAGTCCCAAGAATACTTCTGGAACATAAGCGCATTTTGTACATGGAGTCCCTTGGCCGACATACTTTTTTCTGCGGAAGCAATATAC	216 66720
	N S T L R I H V P L F G S V W P I R Y W A T G S V L L T S D S G R V E V N I G V CAACTCAACGTTGAGATACAGCTTCCCTTTTGGGTGGGTATGGCAATTCGATAGTGGGCCACCGGTTGGTCTCTACAAGCGACTCGGGCTGTGGAAGTAATATGGGT	256 66840
	G F W S S L I S L S S G P P I E L I V V P H T V K L N A V T S D T T M F O L N P AGGATTTATGAGCTCGCTATTCTTATCTCTGGACACCGATAGAATAATGTTGTACCACATACAGTAACTGAACGCGGTACAAGCGACACCATGTTCCAGCTAAATCC	296 66960
	P G P D P G P S Y R V Y L L G R G L D M N F S K H A T V D I C A Y P E E S L D Y ACCGGTCGGATCCGGGCCATCTTATCGAGTTTATTACTTGGACGTTGGATGATGAATTTTCAAAGCATGCTACGGTCGATATATGCGCATATCCGAAGAGGTTGGATTA	336 67080
	R Y H L S M A H T E A L R M T T K A D Q H D I N E E S Y Y H I A A R I A T S I F CCGCTATCATTATCCATGGCCACACGGAGGCTCTGCGGATGACAACGAGCGGATCAACATGACATAAACGAGGAAAGCTATTACCATATGCGCGAAGAATAGCCACATCAATTTT	376 67200
	A L S E M G R T T E Y F L L D E I V D V Q Y O L K F L N Y I L M R I G A G A H P TGGTTTGGAAATGGCCGTACCAGAAATTTTCTGTTAGATGAGATCGTAGATGTCAGTATCAATTAATTTCTTAAATACATTTTAAATGCGGATAGGAGGAGGCTATCC	416 67320
	N T I S G T S D L I F A D P S Q L H D E L S L L F G Q V K P A N V D Y F I S Y D CAACACTATATCGGAACCTCGGATCTGATCTTGGCGATCCATCGAGCTTACGACGAATTTCACTCTTTTGGTCAGGTAACCCGCAATATGCGATTATTTATTTATATGA	456 67440
	E A R D Q L K T A Y A L S R G Q D H V N A L S L A R R V I N S I Y K G L L Y K Q TGAAGCCGCTGATCACTAAAGACCGATACCGGCTTCCCGTGGTCAAGACCATGTGAATGCACTTCTCTCGCCAGGCGTTTATAATGAGCATATACAAGGGGCTGCTTGAAGCA	496 67560
	N L N A T E R Q A L F F A S M I L L N F R E G L E N S S R Y L D G R T T L L L M AAATTTAAATGCTACAGAGGCGAGCTTATTTTGGCTCAATGATTTTATTAATTTCCCGGAAGGACTAGAAAATTCATCTCGGGATTAGAGCGTCCGCAACTTGTCTTTAAT	536 67680
	T S M C T A A H A T Q A A L N I Q E G L A Y L N P S K H M F T I P N V Y S P C M GACATCCATGTTACGGCAGCTACGCGACGCAAGCAGCACTTAACATACAAGAAGGCTGGCATACTTAAATCCTTCAAACACATGTTTACATACCAAGCTATACAGCTCTGTAT	576 67800
	G S L R T D L T E E I H V M N L L S A I P T R P G L N E V L H T Q L D E S E I F GGGTCCCTCTGACAGACCTACGGAAGAGATTCATGTTATGAATCTCTGTCGGCAATACCAACACGCGGAGCTTACGAGGATTGCAATACCACTAGAGGAATCGAAATATT	616 67920
	D A A F K T M M I F T T M T A K D L H I L H T H V P E V F T C Q D A A A R N G E CGACGGGCTTTAAACCATGATGATTTTACCACATGGAGTCCAAAGATTGATATACTCCACACCATATGACAGAGTATTACGTGTCAAGATGACGCGGCGGTACCGAGA	656 68040
	Y V L I L P A V Q G H S Y V I T R N K P O R G L V Y S L A D V D V Y N P I S V V ATATGCTCTATTCTCCAGCTGTCAGGAGACAGTTATGTTATACAGAAACAACCTCAAAGGGGTTGGTATATCCCTGGCAGATGTTGATATATACCCCATATCCGTTGT	696 68160
	Y L S R D T C V S E H G V I E T V A L P H P D N L K E C L Y C G S V F L R Y L T TTATTTAAGCAGGGAATCTGCGTGTGAACATGGTGTATAGAGACGGTCCGACTGCCCATCGGACAAATTTAAAGAAATGTTGATTTGCGGAAGTGTTTTCTTAGGTATCTAAC	736 68280
	T G A I M D I I I I D S K D T E R Q L A A M G N S T I P P F N P D M H G D D S K CACGGGGCGATTATGGATATAATTATTATGACAGCAAGATACAGAACGACAATAGCGCTATGGGAACTCCACAATCCACCTTCAATCCAGACATGCACGGGATGACTCTAA	776 68400
	A V L L F P N G T V V T L L G F E R R Q A I R M S G Q Y L G A S L G G A F L A V GGCTGTGTTGTTTCCAAACGGAATCTGTAACGCTTCTAGGATTCGAACGACGACAGCCATACGAATGTCGGGACAAATCTTGGGGCTCTTTAGGAGGGGCTTCTGGCGGT	816 68520
	Y G F G I I G W M L C G N S R L R E Y N K I P L T - AGTGGGGTTTGGTATTATCGGATGGATGTTATGTTGAATTTCCCGCTTCGAGAATAATAAATAACCTCTGACATAAAAAACATGATAATAAAAACTACTATAAACGATTCTCTA	841 68640
38	CAATACTTTATCGGCAATAATACACACTACCTTTGGGTTTTTCCCGTCCCAATGGTGTGGTGCACCTACCAAAAAATAGAGCGCTAAATATGCTATATAACGCTCCACG - R Q T K K G D G F P T Q H V R G F F L A G L Y A I Y R R G A	8760 512
	AAAAACGGTTCAAAGGCATTACCGATATTGTTAGTACAGGCGAATGGGAATTGATGATCCCAATAACCGCATAGACGCGACAGCGCGCTTATAGCAGGGGCTCCAGAGTACAG F Y P E F A N G S I T N Y Y L A I P I S S G L L R C L R V A G N Y C P D G S Y L	68880 472
	GGTACTAAGTACGGGATATCTCACTATGCTTCCGTGACAGAAACATCAACCGGAACAGTATCCGATAAACCACTCTGTTTTGCAAGGCGTAAATTCGACACCTCTCTT T D L Y R S I E Y E H R E T V S V D V P V T D S L G V G T K A L R L I R V G E K	69000 432

1785

[illegible]

Q L L G V L L E K A P P L S L L S P I N K F Q P E G H L N R V A R A A L L S D L	234
TCAACTCTTGGTGTCTATTGGAGAAAGCCCCACGCGTATCGCTGCTTCCACCAATTAATAAATCCCAACCGGAGGACATCAAACTGTTGTCACGCGCGGCCCTACTTCGGACCT	72240
K R R V C A D M F F M T R H A R E P R L I S A Y L S D M V S C T Q P S V M V S R	274
CAAACGTAGAGTCTGTGCGGATATGTTTTATGACCCGACACGCCAGGGAACATAGGCTGATCTCTGGGTATCTGTGCGGATATGGTTTCGTGACCCCAACCATCGGATGAGTATCAGC	72360
I T H Y N T R G R Q V D G V L Y T T A T L K R Q L L Q G I L Q I D D T A A D V P	314
AATAACTCATACAACACTCGCGGACGCGAGGTTGACGGTGTGTGGTAACAACAGCAACCTTAAACAGGCAACTATTACAGGGAATTTTACAATTTGACGACACCGCGCTGACGTACC	72480
V T Y G E M V L Q G T N L V T A L V M G K A V R G M D D V A R H L L D I T D P N	354
AGTAACATATGGCGAAATGGTTCTACAGGGGACAAACTTGGTAACGCCCTTGTGATGGGAAAGCGCTGCGCGGAATGGATGATGAGCCGCCCATCTCCTTGATATAACGACACCTAA	72600
T L N I P S I P P Q S N S D S T T A G L P V N A R V P A D L V I V G D K L V F L	39
CACGTTAAACATACCGCTCTAATCCCCCAACATCCAACTCCGATTCACGACAGCTGGCGTTCGCGTTAACCGCGGTGTCCTGCGGATTTAGTGATGTTGGGGGATAAACTGTATTCTY	72720
E A L E R R V Y Q A T R V A Y P L I G N I D I T F I M P M G V F Q A N S M D R Y	434
AGAAGCATTAGAAGCGCGGCTACCAAGCTACGCGGTGCTACCTCTCTATTGGAATATAGATATTACGTTTATCATGCCAATGGGAGTGTTTCAGGCAAACTCCATGGACAGATA	72840
T R H A G O F S T V S E Q D P R Q F P P Q G I F F Y N K D G I L T Q L T L R D A	474
TACACGACACGCCCGGATTTTTCACTGTATCCGAACAGGATCCACGCTCAATTTCCACCCCAAGGAGTTTTTTTTATAAAGATGGGATATTACACAGTTGACTCTTCGTGATGC	72960
M G T I C H S S L L D V E A T L V A L R Q Q H L D R Q C Y F G V Y V A E G T E D	514
AATGGGATCCATCTGCCACAGTTCATTGCTGTATGTCAGGCCACACTTGTGGCTTCCGCCAACACATTTAGATCGTCAGTGTTATTTGGTGATACGTGGCCGAGGGTACAGAGGA	73080
T L D V Q M G R F M E T W A D M M P H H P H M V N E H L T I L Q F I A P S N P R	554
CACATTGGATGTTCAAAATGGGAGGTTTATGAAACGTGGCAGATATGATGCTCTCATCACCCTCATTTGGGTAACGAACTTTAAACAATTACAGCTTATAGCTCCGAGCAACCCGCG	73200
L R F E L N P A F D F F V A P G D V D L P G P Q R P P E A M P T Y N A T L R I I	594
TCTAAGGTTGAATTAACCCCGCTTGTATTTTTTGTGTCACCGGGGACGTAGACCTTCCCGAGCCGAGCGTCCCGGGAAGCCATGCCAACCGTTAACGCAACATTACGGATTAT	73320
N G N I P V P L C P I S F R D C R G T Q L G L G R H T M T P A T I K A V K D T F	634
CAACGGAAACATCCCGTGCTCTATGTCCCATTTCAATTTCGAGACTGTCGCGGAACCAACTCGGTTGGGAAGACATACAATGACCCCGGCAACCATTAAGAGCGTAAGGATACATT	73440
E D R A Y P T I F Y M L E A V I N G N E R N F C A L L R L L T Q C I R G Y M E Q	674
TGAAGACCGCGCATACCCAACATTTTCTACATGCTAGAGGCTGTTATTCATGGAAGCAAGAAAGAACTCTGTCGCGTACTCGGCACTGTTAACACAGTGATTTCGCGGGATTGGGAGCA	73560
S H R V A F V N N F H M L M Y I T T Y L G N G E L P E V C I N I Y R D L L Q H V	714
ATCCCCACGGGTGGCATTGTGAATAAATTTTCACATGTTAATGTACATAACTACATATCTCGGAACGGTAGGCTTCCCGAAGTCTGATTAAATATATCGGGATTACTGACAGATGT	73680
R A L R Q T I T D F T I Q G E G H N G E T S E A L N N I L T D D T F I A P I L M	754
AAGAGCATTACGCCAAACTATAACCGATTTTACAATACAAGGAGAGGGCCATAACCGGAGACCTCGGAAGCGCTAAATAACATCCTTACGGATGACACGTTTTATTGCACTTATCTATG	73800
D C D A L I Y R D E A A R D R L P A I R V S G R N G Y Q A L H F V D M A G H N F	794
GGATTGTGATGCGTTAATATACCGTGATGAAGCGCCCGAGACGACTCCCGCAATTCGTGTGAAGCGGGGAAACGGATACCAAGCCCTTCACTTTGTGGATATGGCCGGGACATACTT	73920
Q R R D N V L I N G R P V R G D T G G Q G I P I T P H D R E N G I L S K I Y Y Y	834
CCAACGACGGGATAATGTGTTAATCCAGGGAGACCGGTGCGGGAGACACGGGTACGGGATTECCCACTTACTCCACACCATGACCGTGAATGGGGTATTCCTCCAAGATTACTACTA	74040
I V I P A F S R G S C C T M G V R Y D R L Y P A L Q A V I V P E I P A D E E A P	874
TATTGTCACTTCGATTTTCCCGCGGTCCTGTGTGACATGGGCGTGGTTATGATGCGCTATACCTCGCTTACAGGCAAGTTATGCTTCCGGAAATTCCTCGTGATGAAGAAGCCCC	74160
T T P E D P R H P L H A H Q L V P N S L N V Y F H N A R L T V D G D A L L T L Q	914
AACATCCCGAGAAGATCAAGACACCCCTCTTCAGCGACCAACCTGTTCCGAACTCTCTTAACGTTTACTTCCATATGACACACTAACCGTGTGATGGATGATGCTGCACACTACA	74280
E L M G D M A E R T T A I L V S S A P D A G A A T A A T T R N M R I Y D G A L Y H	954
AGAGTTAATGGGAGATATGCGTGAAAGCAAGCAGGCCCATTTAGTATCAAGCGCCCCGATGCGGGAGCGCCACGGCAACCAACAGAAATATGAGAATATATGACGGAGCGCTTACCA	74400
G L I M M A Y Q A Y D E T I A T G T F F Y P V P V N P L F A C P E H L A S L R G	994
TGGCTTATTATGATGGCATATCAAGCGGTACGATGAACCAATGCAACGGGTACTTTTTTTATCCGTCCTGGGTCAACCTCTGTTTGCATGTCGGGAACATTGGGCATCATTCGCTGG	74520
M T N A R R V L A K M V P P I P P F L G A N H H A T I R Q P V A Y H V T H S K S	1034
AATGACAAATGCTAGCGGGTTTTGGCAAAAATGGTACCACCAATCCCTCTTTTCTGGGAGCCACCAACCAACTATACGGCAACCCGTTGCCATACATGTAACGATAGTAAGTC	74640
D F N T L T Y S L L G G Y F K F T P I S L T H Q L R T G F H P G I A F T V Y V R Q	1074
GGATTTTAATCTCTACATATTTCTCTTCTGGAGGGTATTTTAAGGTTTACCAATATCTCTTACACATCAACTACGAAACGGATTTCACCCCGGGATGGCTTTACCGTAGTGCGCCA	74760
D R F A T E Q L L Y A E R A S E S Y F V G G Q I Q V H H H D A I G G V N F T L T Q	1114
GGATCGCTTTGCCACAGAGCAACTTTTATAGCCGAGCGTGGCTTGAACTGCTATTTGTGGCAAAATCCAAGTACACCATCATGATGCTATTGGGGGGGTAACATTTACCTTAACCCA	74880
P R A H V D L G V G Y T A V C A T A A L R L T D M G N T A Q N L F F S R G G	1154
ACCCAGAGCTACGTTGGACGCGGAGTGGGATACAGCTGTATGTGCCACAGCAGCCCTGGCATGCCCTCTACGGATATGGGCAATATGCCCAAAATCTTTTTTTTTCACGAGGAGG	75000
V P M L H D N V T E S L R I T A S G G R L N P T E P L I F G G L R P A T S A	1194
AGTGGCAATGTTACATGATAACGTTACCAAGTCGTTGCGTGTATAACAGCATCGGGGGGTCGTTAAATCCCAACGAAACCCCTACCATCTCTCGCGGAGTACGCTGCTGATACATCGG	75120

VZV DNA sequence

1787

GIARGQASVCE FVAMPVSTDLQYFRTACNPRGRASGMLYM 1234
AGGAATTGCAGAGGCAAGCCTCTGTGTGTGTTGTGGCCATGCCGTGTCCACTGACCTACAATATTTAGAACTGCATGCAATCCTAGAGGTCGAGCATCTGGAATGTTATATAT 75240

GDRDADIEA IMF DHTQSDVAYTDRATLNPWASQKHSYGOR 1274
GGGTGACCGTGACCCGACATAGAGGCTATAATGTTGATCACACAAATCGGATGTTGCTTATACAGATCGAGCACTCTTAACCCATGGGCATCAAAAACATTCATACGGTGACAG 75360

LYNGTYNLTGASPIYSPCFKFFTPAEVNTNCHTLDRLLME 1314
GCTATACAAACGGAACATACACCTTACAGGCGCTCTCTCTATCTACAGCCCATGCTTAAAGTTTTTACACCAAGCGAGGTTAACCTAATTGTAATACACTGGATCGGCTTCTAATGGA 75480

AKAVASQSSSTDTEYQFKRPPGSTE MTQDPCGLFQEAYPPL 1354
GGCAAGGCTGTGGCTGCAAGCTCCACCGACACTGAAATCAATTTAAACGCCCTCCCGGTTTACCGAAATGACACAGGATCCGTGTGGCTTTTCAAGAAGCATATCCACCACT 75600

CSSDAAMLRTAHAGETGADEVHLAQYLRDASPLRGCLPL 1394
ATGCTAAGCGATGCGGCATGTTACGAACGGCTCACCGGGAGAAACCGGGGAGATGAAGTTCATAGCCCAATATCTGATTCGAGACGCGTCCGCCCTTAGGGGATGTTCTCTCT 75720

PR - 1396
TCCGGGATAATTTACCACGCGCCACATACCCACTCCCAATAAAGCCCTGTAGAGCGATTGGCATCTTACTTGAGATTGGATACGCTCGGCGACTTGGTCTGTTTACGCTTCTCTTA 75840

41 NAMPFEIEVLLPGE LSPAETSALOKCEGKIITFS LTRH 38
AACACATGGCTATGCCATTTGAGATAGAGATTGTTACCAAGGAACTATCCCGCGGAAATCTGCTTACAGAAATGTAGGGAAAAATTTACCTTCTCAACCCGCTGCAT 75960

RASLV DIALSSYYINGAPPDTLSLLEAYRMRFAAVITRVI 78
CGAGCTTCACTGGTGGATATAGCGCTGCTCATATTACATTAACGGTGCCTCCACGACACGCTCTCGCTGTTAGAGGCATACCGAATGCGATTGCGGCAGTTATAACACGGGTCATC 76080

PGKLLAHAIGVGTPTPGLFIQNTSPVDLCNGDYICLLPPV 118
CCGGGAAAGTGTGGCGCATGCCATTGGCGTGGGATCTCTACACCGGGTGTGTTTATCAAAATACATCCCCGTTGATCTTGTAAATGGCGATTACATCTGCTTCTCTCGGTT 76200

FGSADSI RLD SVGLEIVFPLTIPQTL MREII AKVVARAVE 158
TTCGGGTCGCGAGCTCAATTCGCTTGGACTCTGTAGGACTGGAATGTTTCCCTTTAACTATCCCCAGACCTTAATGCGAGAAATCATCGCAAGTGGTGCACGGGCGCTTGAG 76320

RTAAGA QILPHEVLRGADVICYNGRRYELETNLQHRDGS D 198
CGACGGCGCGGGTCTCAAAATTTACCCACGAGTTCACGAGGCGGATGTCATTGTTCAATGGAAGCGGTTATGAATCGAAACAAATTTACAACATCGGGACGGATCGGAT 76440

AAIRTLV LNL MFS INEGCLLL LAL IPTLLVQGAH DGYVNL 238
GGCGCTATTCGCACATTGGTTTTAAATCTAATGTTTCCATAACGAGGATGTCTGCTTTTATGGCGCTGATTCACCTTGTGTAGTCCAAGGACACACGCGTTATGTAATTTA 76560

LIQTAN CVRETGQLINIPMPRIQDGHRRFP IYETISSWI 278
TGTATACAAACGGCAATTCGTTAGAGAAACCGCGGTTAATTAATACCGCAATGCCCGGATTCAAGACGGCATCGCGGATTCCCATATATGAAACTATTTCATCTTGGA 76680

SYSSRLG DTYLGT RAILRVCFVFDGPSTVHPGDRTAVIQV - 316
TCAACATCATCTAGACTGGGGATACCTTGGGAACTCGCGCAATTTACGCTCTGTGTGTTGATGGACCTCTACTGTTATCGGGAGACCGACGCGCGGATTCAGGTGAACA 76800

GGTGTAAATAAACAACACAGCTAGTTACATTTACGCGCTCTGTTTTTATTAAGGCATAAACCGGAATCCGGTATACATGAATGCAATATACAGGACATAATTAATGCA 76920
* - KIPMFVSDPICSSGIYVSMILA 374

42 ACCATGAGTATCTGACATTTGTCCTGGTACCTTTACCGGTGAAGTTTTGTGCTAGATTACCAACCGCTTTAATACCTCTGTGAGGTTATCCAATGTTTACATAGATAC 77040
VMLDDSM TGTGTGKT YTKTDLNGMGGKIVETLNDLQKCLY 334

TCCACGGGCTACACCTAACTTTACTGTAGGATACAAGCTCTGTGAGGCTATTATATTTCCGGAGTTAAATCGTTAACAAAATAGTCTACGGCGGCTTTTGTGTTTGTAT 77160
EVPDVG LKVTLSVLEQSAIINGSNFRKVFYDVAPT KQKQL 294

AAAAAAGGGTACGCCACGCTACATCCGGGAGGTATGGAATGATAAACAGTAACACTGGAGCGGAAGATAGCACGTTTCCCTTTTCGAGGACAGCAAACTGTTGTCTATAGCCAAC 77280
LFFPYAVS CGPPI SHYFLLVPASSLVNGKELVAFQQAIAL 254

GATATGCAACTGCAGAACTCGGCTGCTGTTCCCTCTATAGAAACGTTACGTTGTAATGTATTGGGGGTAAAGCGAGTATGTGGCTAAGCATTGAGTAACGCAACGCCCTATC 77400
SIAVASDQSSNGEISVHVNTFTNPHLALIHGLCQTVCRGI 214

TCATGGAAGAGCTGCCAGTTAAGCTCTAAGAAAAAGTGTCCAATCAAAATATACTCAATCCGACTTATAACGACCAACATCGCTACACCAAGTACAGACGCTGCTGATTGAG 77520
ESSSTGTLARLFFHELGFIIWDSKYRGVIAVGTGSARTNS 174

GTAATGCAAGGCTTACGTAACGTAACACTGACGATAATATAGCAAACTGCAACGGTTGACGGCGGATATAAATAAACCTCTACGGGAGTTTTGTAAATAATGGCGGTC 77640
TFAPD VVYLVSSLIACNAVTS PRYLIFRERATKTFLPRD 134

AACCCACACCCAGAAATCTGTTACGCCACCTACAATTTCTGCACGAAGGAGTGGCCATAAATACTGAGTGCAGCGCATGGCTCCATCCATTGTGATGAAACCGGCTTA 77760
FGYGGG SNQKRGGVIEQVFS DAWFLDATR RMAGDMTIFVPK 94

TTAATACATAACGGAACAGCTGTGACATCGCTATGTGCTAAACACCGGCGATGTGCTGCTACATATGTAAACAGTTTAACTGATCCGACGATCCAGTAAGTTATAC 77880
NLVYCSCATVDSHALVRPMHDDCVYTVVNL LQDSSGR LNY 54

AAAAAATGTAAGTCTGCTTTCCGGTATTGTGATGAAACAAAATATTTTCAATGTTGATTTAAAAATCCGACTATAGTTGTACAGCATCAGGTCGAATAAATTAAGTTCA 78000
LFSTSAKG TINTSSVFIIKCN TQNLFGVITQVADPRIFNAE 14

TCCCAACAGAGATTAATCTTGAACCTCGGATACCTGGACGATAGAAAGATATAGTTACCCCAACAAAGTTAAATGATCTCTAAATACCAGTACGTAATAAATGTTTGA 78120
DYFLLNF DQGRIG <---- spliced from 45 1

43	MEAHLANETKHALWNHDHTKGLLH	24
	TACGTACATATTTCTTTTTTCCAGTACACCATATCCGGGTGTAATGGAGGCCATTGGCAATGAAACCAACATGCACTTTGGCATAATGATCACACAAAGGATTACTACA	78240
	VVIPNAGLIAAGIDPALLILKKPGRFKVEVOTRYHATGQ	64
	CGTGTGATACCTAACCGGGGCTTATTCGGCGGAATAGATCCCGATTACTGATTTAAAGAAACCGGACACGCTTCAAGGTGAAGTACAAACAGATATCATGCTACAGGTCA	78360
	CEPWCQVFAAYIPDNALTNLLIPKTEPFVSHVFSATHNSG	104
	ATCGGAACCGTGTCAAGTTTCGCCGCTACATTCGGGATAACGCTTAACAAATCTCTTAATACCAAAACCGGAACCAATTTGTTTACACAGTTTTCGGCCACGCAATATTCAGG	78480
	GLILSLPVYLSPLGLFFDAFNVAIRINTGNRKHRRDICIMY	144
	GGGATTGATTTATCATTGCTGTATCTTAGCCCGGTATTCTTTGATGCAATTAACGCTGTAGCGATACGAATAATACGGAACCGCAAGCACCCTGATATTGTTATTATGTA	78600
	AELIPNGTRYFADGQRYVLLCKQLIAYIRCTPRLASSIKI	184
	TGCAGACTAATCCCAACGGAACCGTATTCTTGTGATGGACACGGGTACTTTTATTAAGCAACAGCTGATGCGTATATCCGATGCAACCCCTCGCTTGCATGCTATATAAAT	78720
	YAEHMYAAMGESHTSNGDNIGPVSSIIDLDRQLTSGGIDD	224
	ATACGACAGCATATGGTGGCAGCATGGGTGAATCACACGCTCAATGGGGACAATATGGACCGTTTCATCATAATCGATCTGTATGCAAGTAACTTCGGAGGTATTGATGA	78840
	SPAETRIQENNRDVLLEIKRAVNIIVNSRHPVVRPSSSRVAS	264
	CTCCCTGCTGAAACAGCATACAGGAAATTAATCGGACGCTCTTGAAGTAATAAACCGGCGTAACATGTTTAACTCCAGGACCCCGTCCGACCTTCTAGTTCGCCGTGTCATC	78960
	GLLQSAKGGHGAQTSNTDPIINNGSFGDGVLEPPGGGRFTGKK	304
	TGGTGTCTCAAGTGCAAGGGCCACGGAGCAAACTTCCACACAGATCCGATCAATAACGGTCTCTTGAAGGCTCTTGAAGGCTCTGAGCGCTGAGCAAGGGCAATTCGGGAAGAA	79080
	NSSASIPPLQDVLFTPASTEPQSLMWFIDICYAQLVSG	344
	AAACAATTCGTCGCCAGCATCCACCTTACAAGACGTTCTATTGTTACCCAGCTTCGACAGAACCCCAAGTCTTATGGAATGGTTCGACATCTGTTATGCCAATTAGTTAGCG	79200
	DTPADFNKRRPLSIVPRHYAESPSPLIVVSYNGSSAWGGR	384
	GGACATCCAGCAGATTTCGGAACGGCGTCCCTATCAATTTGATCCGAGCATATCCGAGAACTCCCAAGTCCGTTGATGATGATCTTACACAGGATCTCTGCTGGGGAGGAGC	79320
	ITGSPILYHSAQAIIIDAACINARVDNPSLHVTAQELVA	424
	TATTACGGAGTCCAATTTATATCACTCTGACAGGCTATTATTGATGCTGCTGATATAATGCGGGTTGACAATCCCAAGGCTACATGTGACAGCTCCGCAAGAGCTAGTCCG	79440
	RLPFLANVNLHNQTPLPAFKPGAEMFLNQVFKQACVTSLTQ	464
	GGTTTACCGTTTGGCTAACGCTCCTAAATAATCAACCCCTTACCGCTTAAACACGGCGGAAATGTTTTAAACAGGTATTTAAACAGGCTGTGTGACATCGTAACCCA	79560
	GLITELQTNPTLQQLMEYDIADSSQTVIDEIVARTPDLIQ	504
	AGGCTTTAAGCGAGTTACAAACGACCCGACTTACAAACATCATGGAATATGATATTGACAGATCTTCCCAACGGTTATTGATGAAATGATCCCGCACACCAAGCTGATTCA	79680
	TIVSVLTMSMDAFYNSSLMYAVLAYLSSVYTRPQGGGYI	544
	GACTATGTTTCGGTGTAAACGGAATGTCAATGGATGCGTTTAAACAGCTCTTGTATGATGCGGTTTGGCGTATCTGTCATCTATATACACAGCAAGGTTGGGGGTATAT	79800
	PYLHASFPCLWLGNRSIYLFDDYNSGGGEILKLSKVPVPVAL	584
	ACCTACCTTACAGCTCTCTCCATGCTGGTATGATGCTCTATATATTATTGACTATTATTAATTCAGGAGGGGAAATCTTAAGCTTCCAAGGTCCCGCTCCCGTAGCCTT	79920
	EKVGIGNSTQLRGKFIRSADIVDIGICSKYLPQGQCYAYIC	624
	AGAAAGGTTGGTATGGTAATCCACACACTGAGGGGTAATTTATACGACGCGGATATTGTTGATTTGGAATTTGTTCTAAGTATTTACCGGCAATGTTACCGGTACATTG	80040
	LGFNQQLQSILVLPGGFAACFCITDTLQALPASLIGPIL	664
	TCTAGGATTTAACAGCAATACAATCAATTTAGTTTACCGGGGGATTTCGGCATGTTTTGTTGATTTACCGATACCTACAGGAGCACTACCTGCATGTTAATCGGACCTATTCT	80160
	DRFCFSIPNPHK - *	676
	AGACAGATTCGCTCTCTATTTCCCAACCCCAATAATAATAGTGTCACTATAAAAAACATAACACAGAACTCTTCATATGTAATTTACGTCATTTCCTCCGTTTCCACCCCTCT	80280
44	MELQRIFFLYTATG	14
	TAAAAATAAAATAACCGGTGGGTGGCATTAAACCCCAAGTACCCGGGGCAATCCGCTAGACTGTTTTCTGCTCATGGAATTACACGATATTTCCGCTGTACACCGCTACGG	80400
	AARKLTPEAVQRLCDALTLDMGLWKSILTDPRVKIMRSTA	54
	TGCAGCGCAAAATTAACCCCGAGGAGTTCAGAGACTCTGCGATGATTAACGCTGGATATGGGATTATGGAAGTCCATCCTGACCGATCCCGGGTGAATAATGCGATCACTGC	80520
	FITLRIAPFIPLQTDTTNIAVVVATIYITRPRQNNLPPKT	94
	TTTTATACTTAAGGATCGCTCCGTTTATCCCTTCAACCGGATCTACTAATATTTGCGGTTGTTGAGCCACAATTTACATCACGCGCCCAAGTCAAGTAACTTACCTCCGAAGAC	80640
	FHVIYVFNVEVSYAMTATLRIYPVENIDHVFAGATFKNPJA	134
	TTTCATGTAATTGTAATTTAATACGAGGCTCTGACGCAATGACGGGACTTTAAGAAATTTATCCGTTGAAACATAGACCATGTTTTGGAGCAAGCTTTAAGAACCGGATCGC	80760
	YPLPTSIPDPRADPTPADLTPTPNLSNYLQPPRLPKNPYA	174
	GTACCCCTTCAACATCTATTCCGATCTCGAGAGATCCACCCCGAGATCTTACACCAACGCAAACTTAAGCAACTACTTACACCCCGGCTTCCGAAAAATCATAAGC	80880
	CKVISPGVWNSDERRRRLYVLAMEPNLIGLCPAGWHARILG	214
	ATGTAAGTATTCTCCGGAGTGTGGTGTGACGACGAAGGCGTTTATGTAAGTGGTATGTAAGGCTATGTTCCCGGATGCGATGCTCGGATCTGG	81000
	SVLNRLLSHADGCDENHRVHVHVGALYALPHVTNHAEGCVC	254
	CTCTGATTAATCGACTCTCAGCATCGGACGGATGTGATGAATGAATCATAGAGTTACGTTGGGGGACTGTATGCGTTACCCATGTCACAAATCATCGGAAGGTTGTGTG	81120

VZV DNA sequence

1789

L H 24
TACA 78240G Q 64
GTCA 78360S G 104
CAGG 78480M Y 144
TGTA 78500K I 184
AAAT 78720D D 224
ATGA 78840A S 264
ATC 78960K K 304
AGAA 79080S G 344
GCGG 79200G R 384
AGCG 79320Y A 424
TCGC 79440T Q 464
CCCA 79560I Q 504
TTCA 79680Y I 544
ATAT 79800A L 584
CCTT 79920I C 624
TTTG 80040I L 664
TTCT 80160

CTCT 80280

T G 14
CGGG 80400T A 54
CTGC 80520K T 94
AGAC 80640I A 134
TCGC 80760Y A 174
ACGC 80880L G 214
TTGG 81000V C 254
TGTT 81120

W A P C M W R K A G Q R E L K V E V D I G A T Q V L F V D V Y T C I R I T S T K 294
 TTGGCTCCGTATGTGGAGAAAGGCCGCTACGCGGAATTAAGTGGAGGTAGACATTGGCGCCACGAGGTTCTTTTGTAGATGTACCACCTGCATTGCAATTAAGACTAA 81240

N P R I T A N L G D V I A G T N A S G L S V P V N S S G W Q L Y M F G E T L S R 334
 AAATCCTCGATTACCGAAATCTTGGCGAGCTTATAGCGGAACCAACGCCAGTGGTCTCTGTACAGTAAATTCATCTGGTGGCAGCTTTATATGTTGGAGAAACATTAAGCG 81360

A I I N G C G L L Q R I C F P E T Q R L S G E P E P T T T - 363
 GGCTATTATTAACGGCTGTGGTCTGCTTCAGCGAATTTGCTCCCCGAGACACAAGATTATCGGGTGAACCGGAACCTACAACCACCTAGTATACCTTAACCTAACCGCCGTTGTGGA 81480

AGGTATATGTCAACATTTACAGTAATATTAAGGTTAAATTTATAAACTACACGTTGTGTGTGACTTGACGCGAACCCGCTGTGCTGAAGACCCGTCGGTAAATGAAACGT 81600
 spliced to 42 ----> M T N H S S A F V A T S Y S G D T F S F T 332

AATAGATTCGCTTTTACATGATCCAGTAATTTGGCCCAACCACTGTTCCAGCGAGACTTGATACCTCAACACGCGGTTCCGTTGCTTTGCTATATGAGCCGTATAACCCACTTT 81720
 I S E G K V H D V Y N A G F W Q E L R S K I G E F V P E T A K R I H A T Y G V K 292

AATCCTCTAAACGTGGCATTACTAAGCTATTAATGGTACAGAAACCATGTTTCCCATGCTACGTGGTACCAAAACACAGTGTATTTGTTGAAGTGTCTAAACACTGTC 81840
 I G R F T A M V L A I L P V L F W T K G H R R P V L F V T S K Q K F H E L V S D 252

AGAACTTGGCGTGTAAACACTGTACGAGAAACGAGTCAACTCTGTCGGCATGATCGCCCAATAGCACCCGATGAATAAATGCGTGGTGTGATGAGGATCAATTTTGAACAG 81960
 S V S P T N F V T R L F C D V R D A H D G L L V S S I F H T T H M L I M K Q F L 212

TTCCACGCTCCCTTATATCTGCCATAGATTGGAACGTCAACCTTTGCGCGTTTGCCATGACTTCCACACTCTTCAATCTCTCAAAAGATGTTCCACAGGTACGAAACCGTTGTGT 82080
 E L T G K Y R G Y I P V D V K A R K G H S G C E E I S E F S T E V L Y S F R Q T 172

AAAGGTAGACAACTGACAGAACTATCCGACAGAGAAACGCGGAAATGTGTCTATAACCCGCTATACGATTTGATGAGGTGCTGCTTCTCCGGTGAATATTCATAAACTGTAC 82200
 F T S L Q C F S D S L S F A R F T N M V G S Y A N R H P A A E E P S Y E Y F Q V 132

ACTACTGACGCTTTTAAATCAGGGCTTACGTTTGCATTACCGAATATCGCATGGTGGTCAAACTACATTGGGGTACAGTTGTACCTGTTGACGATAGAAACCGCCAAACAT 82320
 S S V A K K L E P S Y N A N V S Y R M P K L V V N P T C N Y G T S S L F A G F M 92

TGCGCGTAGCAGTAGCGAGAACAGTGAATATATCAACAGTGTGAAGCGTCCAAATCCGGAATAACGGCTGATGACGTGGGTACATCTATAGCAAAATTCAGAAACGG 82440
 A R R A T A S F L P I Y E C C N H L T G I G P I V A Q H R R T V D I A F N L F P 52

GATTGGGTTGCGTTTCCAGAGACCTTGCCCGTGGAAACGCGGTAGGGGACTCCAACGTCCCAAGCGTTTATCCCTACGACGCTTTAGACGTTCAAAATATCTTACAGATTCTTC 82560
 I Q T A N G L S G Q R T S C P T P S E L T G F R E D R R K L R E F Y R S E E 12

ACCAAGCGTACGACCAACATTATCAATGACATTAAATCAATTCACGGAATCCGCTCATCTCTGTGAAGCAGTAAACAGGAAGCGCGTATCTTACGTACTGTTACGTATATAT 82680
 G L T R G F M I L S M 1

M S G H T P T Y A S H R R N R V K L V E A H N R A G L 27
 CATAAACATTTTCAGGGCGCATTCATTCACTTTGGTCTGTGACGGCACACCTCAACCTACGCTTCTCATAGGCGTAACCGTGTCAAACTAGTTGAGGCGCATAACCGCGCGGGTTAT 82800

F K E R Y L D L I R G G A S V Q D P A F V Y A F T A A K E A C A D L N H Q L R S 67
 TTAAGAACGAGCTTCGATCTAATCCGTGGGGTGGAGTGACAGATAGTACGACATTGTGTATGCTTACTGTCGAAAAGAGGCTGCGCCGATTAAATAACGACGCTCCGCTG 82920

A A R I A S V E Q K I R D I Q S K V E E Q T S I Q Q I L N T N R R Y I A P D F I 107
 CAGCTCGCATAGCTTCAAGTGAACAGAGATTCGTATATCAATCAAGGTGAGGAACAACAGTATTCACAGATTTTAAATACAAACAGACGCTATATAGCACCGGATTTTATTC 83040

R G L D K T E D D N T D N I D R L E D A V G P N I E H E N H T W F G E D D E A L 147
 GCGGTTTGGAATAACAGAGAGCATATAACGATAATATAGACAGATGGAAGACGCGGTAGGACCGAACATCGAACACGAAATCATCTTGGTTTGGAAGAGACGAGAGCGTTAC 83160

M D A D D T P P N L Q I S P T A G P L R S H H N T D G H E P N A T A A D Q Q 38
 L T Q W M L T T H P P T S K Y L Q L Q L C V P T T I P T D M N Q M Q P P I S 187
 TTACACAATGGATGCTGACGACACACCCCCAACCTCCAATATCTCAACTGACGAGCTTTGCGTCCACCACAAATACCGACGGACATGAACAAATGCAACCGACGCGATCAGCA 83280

E R E S T N P T H G C V N H P N A N P S T A T C M E S P E R S Q Q T S L F L L K 78
 K N E N P P T P H T D V - 199
 AGAACGAGAAATCCAAACCCCCACACGGATGTGTAATCATCTACGCGCAATCCGCTCACTGCAACATGATGGAATCACCAGAACGATCACAACAGACAGCTATTTTATTAAA 83400

H G L T R D P I H Q R E R V D V F P Q F N K P P W V F R I S K L S R L I V P I F 118
 GCACGGCTTAACGAGAGATCAATACATCAACGGAAGGGTGGAGCTTTTCCACAATTAACAAACCCCCATGGGTTTTAGAAATTTCCAATTTACCGTTTAAATGTACCCACTTT 83520

T L N E Q L C F S K L Q I R D R P R F A G R G T Y G R V H I Y P S S K I A V K T 158
 CACGCTCAATGAACGATGTGTTTCTAAATACAGATTGAGATAGACCAAGGTTTGGCGGACGGGAAAGCTATGCGGCTGTTTATATATACCATGCTCAAAAATAGCTGAAAAAC 83640

M D S R V F N R E L I N A I L A S E G S I R A G E R L G I S S I V C L L G F S L 198
 CATGGACAGCTGCTTTTTAATAGAGATTAATTAACGCGATTTAGCGAGTGAGGGTCTTACAGAGCAGGGGAAAGGCTAGGTATTTTACGATAGTTTGCCTTTTGGTTTTCGTT 83760

Q T K Q L L F P A Y D M D M D E Y I V R L S R R L T I P D H I D R K I A H V F L 238
 ACAACCAACAGCTACTGTTTCCGGCATACGACATGGATATGAGTGAATACATGTTTCCGCTGTCCAGACGCTTGACAATACCTGATCAGATAGACAGAAAAATGCCATGATTTTT 83880

D L A Q A L T F L N R T C G L T H L D V K C G N I F L N V D N F A S L E I T T A 277
 AGATTGGCTCAACGCTGACGTTTTTAATCGAACGTCGCGCTGACCCACCTAGATGTGAATGTGGCAATATTTTCTTAACGTCGACAACCTTGGCTCGTTGGAAATACCAACG 84000

VIGDYSLVTLMTVSLCTRAIFEVGNPSHPEHVLRLVPRDAS	317	
AGTAATCGGAGACTAGCTAGTAACATTAATACGATTCCCTTTGTACTCGAGCGATATTGAAGTTGGAAATCCATCCACCCGGAGCAGTACTACGCGTACCCCGGATGCATC	84120	CCGTGG R P
QMSFRLVLSHGNTNQPPEILLDYINGTGLTKYTGTLTPORVG	357	ATGCC H A
GCAGATGTCATTTCGTTTGGTGTGAGTCATGGAACAAACCAACCCCTGAAATCTTGGTGTATTATTAATGGAACGGGCTTACTAAATATACGGAACCTTCCCGCAAGAGTTGG	84240	CAGCA L M
LAIDLIALGQALLEVILLGRLPGLPISVHRTPHYHYGH	397	GACCG V A
ACTTGCAGTGTCTTATGCAATGGGCAAGCACTCTAGAAGTTATCCGTAGGACGCTTCCCGGACAACGCCCCATTTACGATACGCGCCGCAATTATCACTACTACGGTCA	84360	TAATA L L
KLSPDLALDITLAYRCVLAPYILPSDIPGDLNYPFIHAGE	437	TCGGT R H
TAAGTTATCACCAGATTGGCGCTGTATACGCTGGCATATCGATGTGTCTGGCGCATATATATCCCATCTGACATCCCCGGGACTTAAATTAATCCCTTTATACGCCGGAGA	84480	TCCAC G G
LMTRISRNSLRRI:IFQCHAVRYGVTHSKLFEGIRIPASLYP	477	AGCCA A L
GCTGAACCCGCTATTCCCGGAATCTTACGCCGATATCCAGTGTACGCGAGTGGTACGGCTAACGCACTCAAGCTTTTCGAAGGCATACGCTTCCGCCCTCATTATACCC	84600	CGAGC S V
48		
WARSGLORIDISPQPAKK	18	L A TCTGG
ATVVTSLLLCHDNSEIRSDHPLLWHDRDMIGST-	509	P V CCCTA
AGCCACTGTTGTACATGTTGTTGTGTACAGTAATCAGAATACGCTCGGATCACCTTTATTATGGCAGATCGGGATTGGATAGGATCGACATAAGCCCCAGCCAGCCAAAAA	84720	F P GTTAA
IARVGGLOHPFVKTDINTINVEHNFIDTLQKTSPPNMDCRG	58	D S CGACA
ATTGCCGTGTGGGAGGTCTACAGCACCTTTTGTAAAAACGGATATTAACACGATTAACGTGAACACCACTTTATAGACACGCTACAGAAGACATACCCGAACATGGACTGTCCGGG	84840	S C TTCCT
MTAGIFIRLSHMYKILTTLESPPNDVITYTTPGSTNALFFKT	98	G F CGGA
ATGACAGCGGATTTTATTTGTTTATCCCATGTATAAATCTAACAACTCTGGAGTCTCCAAATGATGAACCTACACACACCCGGTTCTACCAACGCACTGTTCTTAAAGAGC	84960	E L CGAAC
STQPQEPPEELASKLTQDDIKRILLTIESETRGQGDNAI	138	S T CTCAA
TCCACACACCTCAGGAGCGCGTCCGGAAGAGTATGATCCAAATTAACCAAGACGACATTAACGATTTCTATTAACAATAGAATCGGAGACTCGTGGTACGGGACATGCCATT	85080	I P CATAA
NTLLRRNLITASTLKWSVSGPVIPPQWFYHHMTTDTYGD	178	P I ACCCC
TGGACACTACTAGACGAAATTAATCACCAGTCAACTCTTAATGGAGTGTATCTGGACCCGCTATCCACCTCAGTGGTTTACCACCAATACACTACAGACACATACGGTGTGCG	85200	F P ATTAA
AAMAFGKTNPEAARAIVEALFIDPADIRTPDHLTPEATTK	218	D S CGACA
GCGCAATGGCTTGGAAAAACCAACGACCGCGGACGAGCGATAGTTGAAGCATTGTTATAGATCCGGCTGATATCCGCTACTCTGATCATTTAACCGCAAGGCTACAACTAAG	85320	S T CTCAA
FFNFDMNLTKSPSLLVGTPIRGTYECGLLIDVRYTGLIGAS	258	I P CATAA
TTTTTAAATTTGACATGCTCAATACCAATCTCCAAGTCTCTTGGGTACACCAAGATCGGAACGATGAATGAGGACTTTAATCGAGCTCGAACGGGACTTATAGCGCGCTCG	85440	P I ACCCC
LDVLYVCDRDLPTGLTLNPHPAETDISFFEIKCRAKYLFDPD	298	F P GTTAA
TTGGACGTTCTGTATGTGACAGGACCTTTAACTGGCACCTTAATCCCACTGACAAACCGACATTTCTATTTTGAATTAATGCTGTGCTAAATACCTCTTGTATCCAGAT	85560	D S CGACA
DKNNPLGRITYTTLINRPTMANLRDFLYTIKNPCVSFFGGS	338	S C TTCCT
GACAAAAATACCCGCTCGGTGGGACGTACACACGTTAATAATAGACCTACAAATGCAATCTACGGGACTTTTATATACATAAAAAACCAATGATGAAGCTCTTGGACCTCA	85680	G F CGGA
ANPSTREALITDHVEMKRLGFKGGRALTELDAHHLGLNRT	378	E L CGAAC
GCAACCAAGTACCGGAGGCTTAATAACGGATACGTTGAATGGAAACGTTAGGATTAAGGTTGGAGGGCCCTACAGAACTCGAGCCCATCATTTGGGCTCAATCGGACA	85800	S T CTCAA
ISSRVWVFNODPIDQKGTITTIATWATGDTALQIPVFANPRH	418	I P CATAA
ATCTCATCCGAGTGGGTATTTAATGATCCGACATACAAAAGGGGACAATTACAACTATGATGGGCACTGGAGATACGGCTCTCAAAATCCGTTGATTGCCATCCCGGCAC	85920	P I ACCCC
ANFKQIAVQTYVLSGYFPALKLRPFVLYTFIGRVRRPHEVG	458	F P ATTAA
GCTAATTTAAACAAATGCCGTACAAACCTATGTATTATCCGGTACTTTCCAGCGCTAAACCTACGCGCTCTCTGTCACCTTATAGGAGCTGTGGCGCGACACACGAGGTGGGA	86040	D S CGACA
VPLRVDTQAAAIVEYHNPPTIPPHCAVPVIAVLTPIEVDVP	498	S T CTCAA
GTCCCATTTGCGCTGATACACAAGCGGCTGCCATTACGAATATACTGGCCGACTATCCCAACCCACTGTGGGTTATAGCCGTTCAACGCTATCGAAGTTGATGTGCT	86160	I P CATAA
RYTQILKDTGNNAITSAALRSRLRNDNLHPAVEEESVDCANG	538	P I ACCCC
49		
MGOSSSSSGRGGICGLCKR	18	F P ATTAA
AGAGTGACAAATCTTAAAGACACAGGAACACGCGATTACATCAGCATTGCGGTCATTGCGATGGGACAATCTTATCCAGCGGTCGAGGAGGAATCTGTGGATTGTGAACGGT	86280	D S CGACA
TTSLLRATEKPLL-	551	S T CTCAA
YNELVTCNGETVALNSEFFEDFDENVTEADAKSTORRP	58	I P CATAA
ACAACGAGCTTGTACGTGCAACGGAGAAACCGTGTCTTGAACCTAGAGTTCTTTGAAGACTTTGATGAGATGTAACAGAGGACGCCGATAAATCCACACAACGCCGCCAC	86400	P I ACCCC
RVIDVTPKRKPSGKSSHSKCAKC-	81	F P ATTAA
GAGTGATCGATGAACACCAAAACGAAACCTTCGGGAAAGAGCTCCCATTCCTAAATGCGCAAAATGTTAAACCTGATAACCTGATAACGTTCTAATAAAACATCAATCATGGT	86520	D S CGACA
TGGTACTGTGAATGTTTATGCTTGGGGTTACAAGTAACCCACGCTACTCCCACTGTTGATCGCTGATTAACAGCTCATCTCCGCGGTCCTTTCATATGTTGA	86640	S T CTCAA
50		
-ENGNSNREYLLLEDERDTEYS	415	I P CATAA
GTCATTTTCATAGAGTACGCTAGCCTTGTGATGGGTAATTTGTGGCGGAGAAATTTCTATGTGAGGTTTACTTTTCGTATGTATCCCGTACCCGCTCGGGTACTTCTTACGCG	86760	P I ACCCC
DNEYVYGYGQSPYNTRRSNRHAPKVRIYGRVREPVRVA	375	F P ATTAA
ACCGTAGAACCGCTGCTTCTGTCGATGATACATATGACGACATCAATCTGAGAAGCAACATGACACGGAACACCGCCAGGCAAGCAAGGTTCCCGAGTTGTGGGAATTA	86880	D S CGACA
G YFRSRKQRHYVYARMLRLLLNMYVSFVALCALTGRTTPII	335	S T CTCAA

1791

CGTGGAGATTGAACCGATATAGGGTCATATAATCGGTCCATATACGAGTGCCTGGCGGCGTTCCCAACGTAGCACAGCCACGAGCTTCCACGGAGCGTCTTAATTAACCGTGTATATA	87000
R P S Q V S I P D Y L R D M Y S H A A T G L T A C A V L T G L S P G I L V H I Y	295
ATGCGCCAAAATTAATTCGTACTATAAGATATAACAAGTACACTGACATGCTAAATGTAGACATGGCCACGGACACCGATGACCACAGTCCCGATGTAGATGATTCGCCACCACAAGTTC	87120
N A L I L E S V I L V V S L T S F T S M A V S V S S M L G T H L H N A V V L E	255
CAGCATTAAATGATACAAATAGGATACATATCGCCATCAACCGAGCCATCAAAATCACGAACACTGCGCGCTAGGCGCCCGCAGCGCATATAAAAGACGCTCTGCTGCTGTAATTTGC	87240
L N L S V F L I C I A M L A A M L N V F V A R T P G A L R Y L F V S Q Q R L N A	215
GACCGCTTTTATGTCGTTGCTGCAATTTTCGCGGTCACAAAAATACGTTGTAATATTACACTTGTGCGCAAAATGTCCAGATATAATGTAGACAGCCACCGGATTGCTGTGAAGC	87360
V A K I N T E D L K R G C F Y T T F I V S T A F H G L Y L T A A V G I Q K Y A	175
TAATAATACAAACCGGCTTTAATAACCAATGACAAAAAGCCCCAAAAAGTGTGTGGGATCTACAACATAACCATGCAACACGGAGCTTTCGCCGACACAGTGTGATTTTCGTTTC	87480
L L L V V A N L L N L S L L G W F L T T P D V V L M A V G S S Q R V R Q N K T E	135
TCGGTGTAATACGCGCGCGGTATCATAGTATATACCGCATGGCCATTGCGTTAAAGCGGTGTAGTAAGTAAGTCCACACAGCTATGTGGTTCCAAAAACAAACCGGGGCGCTGTA	87600
R H I I A A T I L T Y V A M A M A T L A T Y T F A V V S H P E L F L V P A S Y	95
TCACCTCTATTTCGGGACCATACCCCCCATCTAGGGTGGCGTTAAATAACTCATATAACACTACGGCAGCATAAAAACAGGGATCCCGGTATATTCAGAAGAGGCGCAATTAACGT	87720
G G R N G S M V G G D L T A N F L E Y D V V A A Y F C P I G T V E S S A I L T	55
AGCGAGGACATTACGCCACCCAAAGTGAACATCATCAGCTGAATTAACCAATTCGCCAATTAAGCGTATTCATTGTATGATCATTAACGCTTCCACCTCGGGTGTGCTGCTGCTGCTACCG	87840
A L L N V A G L T F M N V Q I I W I R N N L T D M Q H D L A E V E P T T T D Y P	15
M S P N T G E S N A A V Y A S S T Q L A R A L Y G G D	27
CGAGACTTTTCAGAACCGCGCCCTCTTTTGTAGTTCCTGCTCCCAACACCGGGGAGAGCACGCCGCCGCTATGCGTCCAGTACACAGCTCGCGCGGCGTTATATGGAGGGGA	87960
S V K E S R P G K K Q T G M	1
L V S W I K H T H P G I S L E L Q L D V P V K L I K P G M S Q T R P V T V V R A	67
TCGGTTGCTGGGATTAACACACCCACCCGGGAATTAGCTTGGAACTGCAATTGGATGTTCCAGTAAAACTAATAAAACCTGGTATGTACAAAACTCGCCCGGAATCCGCTGCTACGTGC	88080
P N G S G K T T A L L E W L Q H A L K A D I S V L V S C R R S F T T Q T L I Q R	107
CCCTATGGCTCTGGTAAAAACAGCCCTGCTTGAATGGCTTCAACACGCGTTAAAGCAGATATTAGCGTACTGGTGTGCTCATGTCGCGGTAGCTTTTACCAGACGCTGATTCAACG	88200
F N D A G L S G F V T Y L T S E T Y I N G F K R L I V Q L E S L H R V S S E A I	147
GTTTAACGATGAGGCTCTCCGGATTGTAACATATTGACATCCGAGACATATATTATGGGTTTAAACGTTTGAATGTGCAACTGAAAGCCATACACCGGTATCCAGCAGAGCTAT	88320
D S Y D V L I L D E V M S V I G Q L Y S P T M R R L S A V D S L L Y R L L N R C	187
CGACAGCTACGAGCTATTAACTAGGATGAGGTAATGTCAGTATTGGACAATTATCTCCCCCAACATGAGACGCTTTCGCCGGTGTGATAGCCATTATTATGCTCTTTTAAATCGCTG	88440
S Q I I A M D A T V N S Q F I D L I S G L R G D E N I H T I V C T Y A G V G F S	227
TTCTCAAAATTATCGGATGGATGCTACAGTAACTCGCAGTTTATTGATTAAATCTCCGGATTGCGTGGAGATGAAACATACACACAATTGTGTATACGCGGGAGTGGGTTCTC	88560
G R T C T I L R D M G I D T L V R V I K R S P E H E D V R T I H Q L R G T F F D	267
CGGAAGAACTTGCAGATCTCGGTGATAGGCACTGACACGCTGTGCGAGTCAATAAACGATCTCTGAAACACGAGGATGACGTACCATACACCAACACGTTGGGAACATTTTTTGA	88680
E L A L R L Q C G H N I C I F S S T L S F S E L V A Q F C A I F T D S I L I L M	307
CGAATAGCACTACGATTACAATTAGTGGGCTAACATCTGTATATTTTCACCACTTATCGTTTTCGGAGCTAGTGTGCTCAGTTTGTGCAATTTTACAGACTCTATTCTATTATTA	88800
S T R P L C N V N E N K H F R V L V Y T T V V T V G L S F D M A H F H S M F A Y	347
CTCAACTCGGCCCCATGTGAATGTAACGAATGGAAACATTTTCGCGTGTGGTGTACACTACCGTGTGACCGTGGATGGATGAGTTTGTACATGGCTCATTTTCAATGATGTTTGTCTTA	88920
I K P M S V G P D M V S V Y Q S L G R V R L L L N E V L M Y V D G S R T R C G	387
CATAAGCCAAATGTCATATGGCCGGATATGGTATGGCTTACCAGTCATTAGGGCGTGTAGCTTATGCTACTTAATGAAGTTTGTATGTCAGTCGATGGCTCAAGGACAGATGCGG	89040
P L F S P M L L N F T I A N K F Q W F P T H Y Q I T N K L C C A F R Q R C A N A	427
ACCCCTGTTCTGCCAATGTTACTAACTTTACCATCGCAATAAAATTTCAATGGTTTCTTACACACACCCAAATAACATAACCAACTGTGCTGTGCTATTAGGCAACGATGTGCAAAATG	89160
F T R S N T H L F S R F K Y K H L F E R C S L W S L A D S I N I L Q T L L A S N	467
ATTTACAGCTCGAACACCCATCTCTCTCAAGATTAAATAACAACACCTTTTCGAGAGATGCTCTCTTGGAGTTTGGCCGATAGCATTAAATCTTCAAAACTCTTTTGGCCTCTAA	89280
Q T L V V L D G M G P I T D V S P V Q F C A F I H D L R H S A M A V A S C M R S	507
CGAAATTTTGGTTGATTGGATGGCATGGGCTCAATAACGGACGTTTCCCCAGTCAAAATTTGTGCAATTATACAGATCTCAGACATAGCCGTAACGCCGTAGCTTCTGTATGCGGTC	89400
L R Q D N D S C L T D F G P S G F M A D N I T A F M E K Y L M E S I N T E E Q I	547
YCTTAGACAGGCAATGACAGCTGCTTACCGATTTTGGCCCTTCGGGATTATGGCCGATAAACAATACCGCGTTTATGGAAAGTATCTTATGGAGTCAATTAATACCGAAGAACAAAT	89520
K Y F K A L A C P I E Q P R L V N T A I L G A C I R I P E A L E A F D V F Q K I	587
TAAGATTTTAAAGCCCTTGCATGCTCAATAGAACAGCTAGACTAGTCAATACGGCAATATGGGCGGTGTATACGAATACCTGAAGCTTGGGAAGCATTGACGATTTTCAAAAAT	89640
Y N H Y A S G W F P V L D K T G E F S I A T I T T A P N L T T H W E L F R R C A	627
ATACAGCACTACGCTTCCGGTGGTTTCCCGTCTGGACAAAACCGGGGAATTTAGCATCGGACTATAACTACCGCCCAAAATTTAAACACACATTTGGGAGCTGTTTGCCTGTTGTGC	89760
Y I A K T L K M N P S T E G C V T Q V L D T D I N T L F N Q H G D S L A Q L I F	667
CTATATTCGAAAAACCTCAAGTGAATCGCTCCACCGAAGGCTGTGTAAACAAAGTTTGGATACGGCAATTAATACACTTTTCAATCAACACCGGGGATTCGCTGGCTCAACTAATATT	89880

	E V M R C N V T D A K I I L N R P V W R T T G F L D G C H N Q C F R P I P T K H	707
	TGAGGTATGCGCTGTACGTTACTGACGCTAAGATTATATTAACCGCCGCTTGGCGAACACCGGATTCTTAGATGGATGCCATAAATGCTCCGTCCTAATCCCTACAAACA	90900
	E Y N I A L F R L I W E Q L F G A R V T K S T Q T F P G S T R V K N L K K K D L	747
	CGAATATACATGCTCTATTTCGTTAATTTGGGAACAATTATTGGCGCCCGTAACATAAAGTACCCAGACCTTTCGGGAAGTACTCGTGTGAAAAACCTAAAAAAGATCT	90120
	E T L L D S I N V D R S A C R T Y R Q L Y N L L M S Q R H S F S Q Q R Y K I T A	787
	AGAACTTACTGATTCAATTAACGTGGATCGTTCTGATGTGCTACCTACCGCAGTTGTATAACCTGCTTATGAGCCAGCGCCATTCTGCTCTCAACAGCGTTACAAAATTACTGC	90240
	P A W A R H V Y F Q A H Q M H L A P H A E A M L Q L A L S E L S P G S W P R I N	827
	CCCCGCTGGGCGCCAGCTGTATTTCAAGCAGATCAATGCACTTGGCCCGCATGCGGAAGCATGCTACAATTAGCGCTATCGGAAGTGTCCCGGGATCGTGCCGCGGATAAA	90360
	G A V N F E S L -	835
	CGGGCGGTAATTTTGAAGTTTATAACCGTTAATACCATATATGGACATCCATAGGGGGGTTACATAAATACTAGCCCTGTACAACACAAAGGCTCTAACAATGCACTGAAC	90480
52	M D A T O I T L V R E S G H I C A A S I Y T S W T Q S G Q L T Q N G L S	36
	CACAACCAAGCTATGGACGCAACGAGATTACCTTGGTTAGAGAAAGCGGACATTTGTGCGCAAGCATATACACATCTGACACAGTCCGGACAATTAACACAGACGCTTTTCC	90600
	Y L Y Y L L C K N S C G K Y V P K F A E I T V Q Q E D L C R Y S R H G G S V S A	76
	GTGTATATACTATTATGCAAAACTCATGTGGGAAATACGTCCTAAGTTTGGCAAAATACCGTACAACAAGAGGATTTATGTCGTACTCCAGGATGGGGGAGTGTTCGCG	90720
	A T F A S I C R A A S S A A L D A M P L E P L G N A D T M R C L H G T A L A T L	116
	GCAAGTTTGGCTATCTGCAAGGGCGGCTCTCGGTGCGTTAGACGCTGCGCCCTGAACCACTGGGTAAACGAGACACCTGGCGTGTCTCCATGGCACTGCCGCACTTTA	90840
	R R V L G F K S F Y S P V T F E T D T N T G L L L K T I P D E H A L N D N T P	156
	CGGCGGTATTAGGGTTAAATCGTTTATTCGCGAGTAACATTGCGAGCTGATACGAATACAGGCTCTCTGTTAAAAACAATCCCGATGAACAGCGTTGAATAATGACCAACGCGCA	90960
	S T G V L R A N F P V A I D V S A V S A C N A H T Q G T S L A Y A R L T A L K S	196
	TCTACCGAGTATTGAGGGTAATTTCCCGTGGCATTGATGTTTACGAGTCAGCGATACGCGCCACGCAAGGTACGCTGCTAGCCCTACGCGCGCTGACCGCACTTAAATCT	91080
	N G D T Q Q Q T P L D V E V I T P K A Y I R R K Y K S T F S P P I E R E G Q T S	236
	AACGGTGACACCCAGCAACAACACCTTTAGACGTGGAGGTAATACACCAAGGCTACATACGTCGGAATATAAGTACGCTTTTCCCGCTATAGAGCGGGAAGCCAAACCTCC	91200
	D L F N L E E R R L V L S G N R A I V V R V L L P C Y F D C L T T D S T V T S S	276
	GATTGTATTACCTTGAAGAACCGCGCTTGGTTCTTAGTGGCAATCGCGCAATTGTTGAAGGTACTTTACCGTGTATTGACGTGTTTAAACAGGATCCACGTTACATCTTC	91320
	L S I L A T Y R L W Y A A A F G K P G V V R P I F A Y L G P E L M P K G E D R D	316
	CTTCAATATTAGCAACATATAGACTGTGGTACGCGGCGGCTTGGAAAAACCGGGTTGCGTCCAATCTTGGCTATTATAGCCCGGAACCTCAATCGAAGGGTGAAGACAGAGAC	91440
	Y F C T V G F P G W T T L R T Q T P A V E S I R T A T E M Y M E T D G L M P V T	356
	TACTTTGTACTGCTGGATTTCGCGGATGGACCACTCTTCGAGACAACTCCAGCGTGAATCTATTCGCAAGGCTACGGAAGATGATACATGGAACGAGTGGGTGTGGCAGTAAC	91560
	G I Q A F H Y L A P W G Q H P P L P P R V Q D L I G Q I P Q D T G H A D A T V N	396
	GGATTTCAGGCTTTTCAATATCTAGCCCCCTGGGACAGCATCCCCCTTACCTCCGCGGTGCAAGATCTTATGGGCAATCCCTCAAGATAGTGGACATGCAAGTGTCAAT	91680
	M D A G R I S T V F K Q P V Q L Q D R N M A K F D F S A F F P T I Y C A M F P M	436
	TGGGACGCGGGCGGATATCTACCGTCTTCAACAGCTGTACACACAGATCGTTGGATGGCAAGTTTGAATTCAGCGCTTTTCCACGATATCTCGCTATGTTCCTCCATG	91800
	H F R L G K I V L A R M R R G M G C L K P A L V S F F G G L R H I L P S I Y K A	476
	CATTTAGATTAGGCAAAATCGTCTGGCTAGATGCGTGAAGAAATGGGTCGTAACCCCGGCTTGGTGTCTTTTGGGGGTTACGGCACATCTCCGAGTATATACAAAGT	91920
	I I F I A N E I S L C V E Q T A L E Q G F A I C T Y I K D G F W G I F T D L H T	516
	ATTATTTTTATAGCAATGAATTAGCCTTTCGCTGAACAAACGCGCTTGGACAGGCTTTGCTATATGTAATATATAAAGATGGATTTGGGGAATCTTACCGATTATACATCG	92040
	R N V C S D Q A R C S A L N L A A T C E R A V T G L L R I Q L G L N F T P A M E	556
	CGCAATGATGTTCAGATCAGGACGTTGTTCCGCTTAAATTTAGCGCCACCTGCGAAGAGGAGTACGCGCTTATTACGAATCAACTAGGCTTAACTTTACACCCGCAATGGA	92160
	P V L R V E G V Y T H A F T W C T T G S W L W N L Q T N T P P D L V G V P M R S	596
	CCGGTACTCCGGTTCAGGGTGTGTACACTACGCAATTACCTGGTGTACACGGGAAGCTGCTGTGGAATTTACAAACAACAGCCTCCGGATTTAGTGGCGTCCATGGCGAAGT	92280
	Q A A R D L K E R L S G L L C T A T K I R E R I Q E N C I W D H V L Y D I W A G	636
	CAGCGCGCGGAGATTAAAGGAGCGTCTTCAGGACTCTATGTACCGCAACAAAATTCGAGAACGATACAGGAAAATGCAATATGGACCATGCTATACGACATATGGCCGGA	92400
	Q Y V E A A R K T Y V D F F E H V F D R R Y T P V Y W S L Q E Q N S E T K A I P	676
	CAAGTTGGAGGCTGCCAGAAAACATACGTCGATTTTTTGAACATGTTTGTATCGCGTTATACCTCGGTATCTGGAAGTCTTCAGGAGCAAAATTCGGAACAAAGCAATACCG	92520
	A S Y L T Y G H M Q D K D Y K P R Q I I M V R N P N P H G P P T V V Y W E L L P	716
	GCATCTTATGACATACGGACACATGCAAGATAAGGATTAAACCAAGACAGATAATATGGTGTGTAATCCCAACCCACATGGACCTCTACTGTGTTTACGGAATGCTACCA	92640
	S C A C I P P I D C A A H L K P L I N T F V T I I N H L L D A H N D F S S P S L	756
	TCGTGCGCTGTATCCCCCATAGACTGCGCTGCTCATCTCAAGCCCCCTATACACAGCTTTGTCATATTATTAACCATCTCTAGATGCTATAATGATTTTTCAAGTCCATCATG	92760
	K F T D D P L A S Y N F L F L -	771
	AAATTTACTGACGATCCCTTGGCTCATATAACTCTCTTTTTATGACAAAAAACACGCGCAACACCCATCTTAAAAATAAGAGTTTATTTACTTTACAACCCGTTGGTAATTT	92880
	- K V V R P S N K	324

TATAC

CTTGC

ACTTA

CCAAA

AATCT

ACGCT

AGTTT

GTGTT

T N

TCCG

CTTTC

TAGGA

GGCG

AAAC

CCTG

ACAT

GCCC

ATAT

TGAG

CGCT

TTAT

GAC

GCA

CGTA

AAC

L

VZV DNA sequence

1793

707 1000	TATACGTTTCAATTAACGAACATTTTTCGGTGTACCATGTCGATTAAACCAAAAAATATACGCTCTCTGATATCCGAATCTCGTAAAGGTCATTAAACAATCCCGGGGTA Y T E F L Q V N K P T V M T R N L M W F Y V S K Q Y E S D R L P G N L L G P P V	93000 284
747 1120	CTTGACCAACCATCTGACAGGGGGGGTTCGGTGGGAGGTCAAAACGCTGACCAACCCACATGAATATATAGCTTTATAATATTGGGGCCGTTCCAGGTGAGGGTTCAGTA Q V V G D P C P P T G H P L D F R Q G V G C S Y I A K I I N P A T G P Q P N L L	93120 244
787 240	ACTTAACAACATATAATGCGGCAATACGCGGGTTTTGTAAGGGGGTTGTTATCAACGACATACATTAGAGTGTAAACCAATAAACTCCCTCATAAAAACGACGATTTTT K V F M Y H P L V R T K T F P N N D V V Y M L T N L L M L V G E Y L F R R N K E	93240 204
827 360	CCAAAGGCTCTATTGACACTCAACGCGTAAAGTATACAGACAATTGTACAAACAGCGATGGAGATGCCCGAGGGGCCAATGCCCTCCAGATACATTAATAACACATAAGGTAA L P G I Q C E V R R L Y V S L Q V F L S P S A G S P G I G E L Y M L I V C L T F	93360 164
935 480	AATCTAGGACATTATCGGGCGGAATAGAGTACGATAGATTAACAGGCGGGAGGACCCACCGTATACCCCTATCTTCAACCGCAGTAAATACGAAAAATAATCCGCGGA D L V N D P R F L T M R Y I L L R P P V G V T Y V R D E V A T L V S F I F G R F	93480 124
36 100	ACGCTGGTGGTAACACACTCCATGTAGTAACGATACAGGACACCTCACTGAATCACCATTCAACACTACTAAACGGTCTCTGGTGTCCGGTTTACGCGCAGTATACACAG A P Q T V C E M Y Y R D C S V E S S D G N L V V L V T E Q H E P K V R L S V V S	93600 84
76 20	AGTTTGCCAAAGCGTGGCTTAACCGGGTACCTCCCGGCGCTCGCATCAAGTCTGGTATTGCTGATCTTAAGATCTTCGATCACGTCGTCACATCAACCCCTCTCCGGTC N A L F R P K L G T V E R A E C V F R P I A Q I R L D E I V D S V D L G E E A R	93720 44
16 40	- S R R D R R E C G V G R R S	756
56 60	GTGTAGTAAGTGTGATCGTTACGCTGCAACCTAAATGCTGGGTATATTTATCCGGACATCCATCGCCATCCCGCGCTCCGGTTTGCCTGAATTTATCCAGTAAGGTCGAA T L L N D I Y V S C G L I S P I N I G S M G D A M G A G G T Q E F K I W Y P R I	93840 4
56 60	T N T L Q R D N R Q L R F H Q T Y K N R V D N R G D G R R R N A R I K D L L T S	716
36 30	TCCGCTGCTATTACCTGTGTACCGTAACCTCTCAGGGGGGTGCTTTCATAAAATGGGATAGGTTTTATATCAACATGCATGATTGGTTATTTATTTATGGGTTCGGGATT R Q M	93960 1
36 30	D A A N V K H V R L R E P P H G K M F H S L N K Y G V H M Y Q N N I K N P E P I	676
36 30	CTTTCGTCATCTCTGTAGGGTCAGGCAACCCAGGAAGGACTGGTGTCTCCGCGGGCCCGTTTTATACCTCTCGCGCAACCTGCATTTCATATAATATCGGATTGGGATAAA R E D D E T P D P L G M S P S P T R R P G R K I V E A R V Q M E Y L I R I Q S L	94080 636
16 10	TAGGACTCTGTTCTCGCTTTTAAAAATAGCTGGCATACTCTCTCTGACCTATGTACCTCGCTTGGAGTACCAAGAATCTAATCGGGTGGCCCGTAATATGAATGAAAAATAC Y S E T R A K K F I A Q C L E E E S R H V E S Q T V L F G L R T A R L I F S F Y	94200 596
16 0	GGCGCACTAGTAATGAGATTGACGATTGAATATGATACAGAAATTTCTGGCTGTGATTATTGTTACCGGGTGAAGCTTAAACAGCGCAAGTCTGTGTTCCATAGCTCAGAC P A V L L S I S A N S Y S V S I E Q G Q N N N V R H L K F C R V L E Q K M L E S	94320 556
6 0	AAACGTTTATATCATCTCCATAAGGGGGATATAACGAGATTGAAACATTTGGCAATATATGCATCATCCCTATTATGCCGTAAGATCTATAACCTCGTGATTAAATCGGCAATA L R K I D D G V P P I Y R S Q F S N A I Y A D D G I I G T L D I V E H N L D A I	94440 516
6 0	CGTGTCTCTGCTGCTTGAATATGTGACCTTTAGATGGCTTTATTTTACCTCTCTCCCGTAACCGTTTACGCTCTCTCTTGAACCTGGAGCTTTCGGTCAGATCGCTGTC R T E E A M T I H S G K S P K I K V R E E R L R K L E G E K F Q L R E T L D S N	94560 476
3 3	ACATCTTGAGACCTCAATGGTTTGAATAATATTACATAACCTCGAGCATCGCTGTATGCTGTTAAACACCGAAGTTTAAACGCACTTGAACGTTTGTGTTCCGGACATT V D K L G E I T K F L N N V Y G E L M G N I S N V V S T K F A S Q V N T T G S M	94680 436
3 3	GGCCCCCGTTAAAGGATTGGTTGGCTTGCACCAACCGGTTGTGATGTGCCACCGATCCATCTCTCCAGATGTGATTGCCGTTTCTTAGATAGGAACGATCGTTTCGGTA A G G N F S Q N A K G F G P Q S T D V S G S G E L I H N G T E E L Y S R V T E T	94800 396
3 3	ATATCTCAACATGTCTCATGTTTTTAAAGTAACTATTAGCTTTAAGTGTACAGCGGGCGATCCAGCGCGTGTGTATCGTCTCGCCCATTAACGATCAACCGACGTGTGCTG I D G V H R M N K L N V I L K V L R S A A S G A R T T D N E G M I R D V A R T S	94920 356
3 3	TGAGATCTATCTTCACTTCGGGACCTATTAACACGGGCAAGGGGCTGATTAAACCTTGGCAGACGGAGCATGTTACGTAATGCATAACAGGCCAACACCTCCCGAGAAAGC H S R D D E N R R G I L V R L P A T N L V Q C V R A H E R L A Y C A L V E G S L	95040 316
3 3	CGCTGAAGGGTGAAGTCAAACTACACCTCCCAACATACAAACGGGCGCCACACGACCAACACTCTCCCTTCAAGCCGTTACATCATCTTGGCATAATTAATCTCGGTTATAA R Q L P S D F V V G E G C V V P P W V V L C E G K N G T V D D K A M I L R R N Y	95160 276
3 3	TTATAATAAGACGCGCTCATATAATCCATAATAGCAACATTTGCATACACTCAACTAGGCTTGTGACAACCGCGCTCTCTGGGCAACGTTGCATCGGCACTTTTAACATCTGG N Y Y L R T R D Y D M I A V N Q M C E V L S T V V A A G R A L Y A D A V K L N Q	95280 236
3 3	GACAGTTCTGCGCTGACCCATATACGATTAATGGTGCAGGGGTTCCATTCTGTCTGATGCTACCTTTTACACGGGCAAACTACACAGGCTATCCAGTCCAGTATTTG S L E A A Q G M Y T N L P A P T G N Q E S R V K R V V P V I G V C A I M D V Y K	95400 196
3 3	GCAAAACGACCTTCCATTAAACCACTGGTATAGAGAACCGGTTATCCACGAGAACTCAAGTAACGATGACTGTAATGTTGACGGCAGGTTTCAAAACCTGATGTGCAAGC A F G V R G N L G S T Y L C G T I G R L F E L L S S Q L T Q R W T E F V Q H A L	95520 156
3 3	CGTACGCTTCTGATTCTCCACATAGCCCAATACGTTCCGCTAGAGCCCGGCGATGCAAGTTACATTGTTGGATGTGGTGTCCCAATCTGCTGCTAGGCTCTCATACCGAGTTGATCC R V A E S E G C L G Y R E A L A G A H L N C Q Q I H H E M D A A L D E Y R T A D	95640 116
3 3	AACGCGTTCATCAAAACGGTGGCTGAACATTGCGGAATTACAGTTCCGTAAGCGGTACAGCGCTATATGCTTGTCCATCGGTATATCCAAAGTCACCGGCTAGGATTTTCGAAAC L A N M L V T A Q V Q R I V T E T S R V A S Y I G G G D T Y G F D G A L I K R F	95760 76

AACATACTTTCGGTGGTGGGTGATTAACATCCAGCCATCTTCCCGGAAATGTACAAAACCTATATCCGGGGCGTACTCATTCCAGTATATATCGAACATGTTCTGTATTGGTCA	95880
LNSQTTPHILNMWGDEEPFTCFGIDPAYENWYIDFMNKYQD	36
55	
TTTGGGTACTTCCATTCAAGCCCTGGTCAATAGAAACAGAACTTGTATCTCTTTTCTTCTACTACCGGAACGTGTTATTAAGAGAGAGCTTATTTCGGCCATTGAAACCACGATGAA	M K 2
NPN SGNL GQDISVSSAIRKEESGSSNNFLSTIEAM	96000
1	
RSISVDSSSPKNVFNPEPTPNFGDDSVYLNFTSMHSIQPIL	42
AAGTCAATTTCTGACAGCTTCTACCCAAAAACGTTTTAATCCAGAGACGCCAATGGATTGATGACAGTGATATTTAACTTACCTCTATGATAGCATTCACCTATCTCT	96120
SRIRELAAITIPKERVPRLCNFKQLLELQAPPENQRHELP	82
CTCAGGATTGAGAACTTCCGCAATTACGATTCACAAAAGACGTTTCCCGGTTGTTGGTTTAAACAGTTACTCGAACTGCAAGCGCTCTGAAATGCAAGGAATGAGCTCC	96240
FSVYLISGNAGSGKSTCIQTLNEAIDCIITGSTRVAAQNV	122
CTTCTCGGTTTATTAATAGCGAAATGCCGGCTCCGAAAAAGACGTTGATCCAAACGCTTACGAGCTATCGATTGATATTACCGGATCCACAGGGTGTCTGCCAAAAATGT	96360
HAKLSTAYASRPINTIFHEFGFRGNHIAQLGRYAYNWT	162
TCATGCTAAGTTATCAACGGCTTATGCGAGTCTCCGATAAACACAATCTTATGAAATTTGGTTTTCGCGAAATCACATTCAGGCTAGCTGGGCGTATGCAATATGAGCTAC	96480
TPPSIEDLQKRDI VVYWEVLIDITKRVFQMGDDGRGGTST	202
GACCCCGCTTCTATTGAGGAGCTGCAAAAAGAGATATGTATCTACTGGAAGTTTAAATGATATAACAAAAGAGTGTTCAAATGGGGGACGAGGTCGCGAGGAACATCGAC	96600
FKTLWAIERLLNKPTGSMSTAFIACGSLPAFTRSNVIVI	242
ATTTAAACCTTGGGCAATTGACGTTGCTTAATAAACCTACAGGCTCAATGTCGGAACCGGTTATCGCATGCGGTTCCCTTCGCGCTTTTACCGGAGCACTTATGTTAT	96720
DEAGLLGRHILTA VVYCNWLLNAIYQSPQYINGRKPVIYC	282
TGATGAAGCAGGATGCTAGGCGTCATATCTCACGCGCTGTTTACTGTTGGTGGCTTTGATGCTATATATCAAGCCCTCAGTACATAAAGCGTGAAGACCGGTCATGATG	96840
VGSPTQTDSLESHFQHDWQRSHVTPSENILTYIICNQLR	322
CGTGGTTTGGCCCAACCACTGACTGTTAGATCTCATTTTCAACATGACATGACGCTTACACGTAACCTCTAGTGAATAATACTCACGTATATACTGCAATCAAACTCTGCG	96960
QYTNISHNWAIFINNKRCD EDDFGNLLKTLEYGLPITEAH	362
TCAATATACTACATCTCACATACTGGAATCTTTAATAACAAACAGATGTCAAGAGGAGGATTTGGAATCTTTTAAACGCTTGAGTACGGCTACCTATTACCGAAGCACA	97080
ARLVDTFVVPASYINNPNANLP GWTRLYSSHKEYSAYMSKL	402
TGCGGCTTGGTCGATACATTTGTTGACCTGATCTATATAACAACTCTGTAATCTTCCGGATGGACGCTCTGTTATCTGTCGATAAGGAGGAGGCGCTATAGTAACT	97200
HAHLKLSKNDHFSVFA LPTYTFIRLTA FDEYRKL TGQPG L	442
ACACGCGATTTAAACTATCGAAAAATGACCATTTTCTGTTGGCTTACCGACTATACATTCACCGGTCACGCAATTTGATGAATACCGCAATTAACGGGACAAACCGGACT	97320
SVEHWIRANSGR LHNYSQSRDHD MGTVKYETHSNRDLIVA	482
TCTGTGAACTTGGATCGGCAAACTCGGTCGTTGCAACAATTTCCAAAGCGAGATCATGACATGGGAACAGTTAAATACGAACACATTCAATCGGCACTTAATGATG	97440
RTDITVVLNSLVVVTTRLRKLVI GFSGTFQSF AKVLRD DS	522
CGTACAGACATCACTTACGTCGTAATAGTCTGATGTTGTAACCAAGACTACGTAAGTATGTTTATGATTCAGTGGTACATTTCAATCGTTGCAAGGTTTACGTCAGGACT	97560
FVKARGEITSIEYAYRFLSNLIFGGLIN FYNFLNKNLHPD	562
CTTTGTGAAGGCTCGAGGAGACATCCATCGAATGCTTACCGGTTCTGTCAACCTAATCTTGGAGGCTGATTAACCTTTTACATTTTGTGTAATAAAACCTACATCCGA	97680
KVSLAYKRLAALTLELLSGTNKAPLHEAAYNGAGAGIDCD	602
TAGGATGCTGATACAAACGTTAGTGCCTTAACCTGGAGTTATTGTCTGGAACAAACAAAGCCCTTACCAAGCAGCGGTTAATGGGGGGGCTGTTATCTTTGCAATTACAAC	97800
GAATSADKAFCTKAPESKYTASIPEDPDDVIFTALNDEV	642
TGGTCAGCTACTTCTGCGGATAAGCCTTCTGCTTACCAAGCCCCGAGTCCAAAGTAACGGCTTCCATACCGAAGACCGGATGATGTAATTTTACGGCACTTAACGACGAGGT	97920
IDLVYCQYEF SYPKSSNEVHAQFLLMKA IYDGRYAILAEL	682
TATTGACTTGGTATCTGCGAGTACGAATTTCTTATCCCAATCATCCAATGAGGTCATGCTTCTGTTAATGAAAGCTATTTACGATGCTGATATGCAATATGACGAGCT	98040
FESSFTTAPFSAYVDN VNFNGSELLIGNVRGGLLSLALQT	722
TTTCAAGACGCTTACAAACCGCCCTTTAGCGGTATGTCGATAATGTTAATTTCAACGGAAGCGAGCTTTTGTGCGCAATGTCGGGGGGGCTGTTATCTTTGCAATTACAAC	98160
DYVTL L GYTFAPVPVFVEELTRKKLYRETTEMLYALHVP L	762
AGATACGTATACCTTTTGGGGTATCTTTTGCACCGTGCAGCTTTGTAGAGGAAGTACCGGAAAAAGCTGACCGGAACTACCGAAATGTTATGCTCTACAGCTACCTCT	98280
MVLODQHG FVSIYNANVCEFTESIEDAELAMATTVDYGLS	802
TATGGCTTTACAGGATCAACATGGGTTTGTGTCATCGTAACGCTAACGATGTAATTTACGAGTCTATAGAGGATGCAAGTTGGCAATGGCCACCACGGTGGATGAGCTTAG	98400
SKLAMTIARSQGLSLEKVAICFTADKLRLNSVYVAMSRTV	842
TCTAACTAGCCATGACAATGACGCTCACAGGCTGAGTTAGAGAGGAGTATCTGTTTACGCGGATAAATGCGCTAAATAGTGTATGTTGCCATGTCGCTACGGT	98520
26	
MKNPQKLAITFLPLVVIPTTYTLCI	24
SSRFLKMN LNPLRERYEKS AEISDHILAALRDPNVHVVY -	881
CTCCTAGGTTCTTAAAAATGAATCTAAACCTCTACGGGAACGATGAAAAATCCGAGAAATAGCGATCACATCTTGGCGCTACGTAACGACGCTGTTGTATTA	98640

VZV DNA sequence

1795

K A L Y K N T H A G L L F S F L G F V L N T P A M S I S G P P T T F I L Y R L H 64
A A G C A T T G T A T A A A A C A C G A T G C G G G C T T G C T T C A T T T C A G G T T T G T C T T A A T A C A C C C G C A T G A G C A T C T C T G G A C C C C A A C G A C G T T Y A T T T A T A G G T T A C A T G 98760

G V R R V L H W T L P D H E Q T L Y A F T G G S R S M A Y K T D A R C D T M S G 104
G G G T A G C G G G T T C T C A C T G G A C T T T A C C G G A T C A G A A C A A C A C T C T A C G C A T T T A C G G G T G G T C A A G A T C A A T G C G G T G A A G A C G G A C G C T C G A T G T A C A A T G A C G G T G 98880

G M I V L Q H T H T V T L L T I D C S T D F S S Y A F T H R D F H L Q D K P H A 144
G T A T G A T C G T C T C A A C A C C C A T A C A G T G A C C C T G C T A A C C A T A G A C T G T T C T A C T G A C T T T T C A T A C G C A T T T A C G C A C C G G A T T C C A C T T A C A G G A A A C C C C A C G C A A 99000

T F A M P F M S W V G S D P T S Q L Y S M V G G V L S V I T E D D L S M C I S I 184
C A T T T G C G A T G C G T T T A T G C C T G G G T C G G T T C G A C C A A C A T C A G C T G T A C A G T A A T G T G G G G G G G T A C T A C C G A A A C G G A A G A T G A C C T A T C C A T G T A T C T C A A T T G 99120

V I Y G L R V N R P D D Q T T P T P T P H Q Y T S Q R R Q P E T M C P S S P Q P 224
T T A T A C G G T T T A C G G G T A A C A G A C C T G A C G A T C A G A C C A C A C A C C A C C C G C A C C A G T A T A C A T C G A A A G C G G A C C T G A A A C C A A C T G C C T C T C A C C A C A C C G 99240

A F F T S D D D V L S L I L R D A A N A - 244
C C T T T T T C A C A T C A G A C A C G A C G T T C T G T T A A T T A C G G G A C C C G C A A C C G G T A A A G A C A G A T T C A A G A C A C A T T A T C C A A C T G A T T A C A T T C A T A C G C G A A A A C G 99360

ACACAAAAATTTATTTAACGGCTTTAATTTGAAGACACCTATCCTCTTAACGTTGATGAGCCTTGACGTTGGGTGCCGCGCTTACCGGTATTATACATAACCGATTACCGTGT 99480
- R Q H A K C T P H R A E G T N Y M V S K G H 50

57
T T A C G G A G T C T G A C C A T T T A C C A G T G T A T G T C T G A A T A C G A C G T T G T G T G C C G A C A A A T T A A C T C G C G T A C A A A T T C T G A T G T T C C C C G G C T G G C A A C C T G G C A T T T C C A
K R C D S M K G T Y T Q L V V N N H G S L I L E R V F K Q H E G P T A V S A N G 10

A A C A C A T T A C G T T C T G A C G T C C A T G A C C G T A T T T C A G T A T T A T G T G G T G G T C A A G T A T T T C C T A T G T A A A G G A C A C G A T C T A A A G C C G T A A A C T A C A C A A A C A C 99720
F V N R E R V D M 1

58
- T R T R G H G S N E T N I P Q D T L T N E K H L L V R D L A T F E Y V F V 185

T G T A C C A C G G A C G A T T T C C G C C G T G A G C G G T G T A A T A C G G A G G T C T T T G C A C G A A T A C T C G T A C A G T A G G T T C T G A C A C G G G T G C A T G G G T T T T T G A C A C A A 99840
P V L P R S K G D T S R T Y V R R P R R A R I S E Y L L N R V R P A H T K Q C L 145

C A C A A A C A T T G C A G G C T T A T G A C T G G A T G G A T T G A A T T T T T A G A T A G G G T C A C G T G T T T G T G T G A C A C G C C T C G A C C A G A A A G G T G C G G T T T T G T A C A C G A C C G T 99960
V F M Q L S K H S S P N F K N K S L T V H K Q R S V G R G S F A A T K T C A V T 105

T A T T T C A C A G G C T T C A T A C C A A G C T G C G G G A T G G T G C G G T A A T T G T C C G C C A A G T T C G T C A A T A G A T G A T A C C A T G A C A C G T A C A A T G T A C A T A G T C G T T T G G T 100080
I E C A N M V L S R R I T D T L Q R R G L E D I S S V M F L T D F P V Y D D K T 65

T T T C T C A A T A C A C G C C G G T G C C A A T C G A A A T T T T C A T T T G C A T C A C G C T A T T T T C T G T A A A T C G T T C T G A A C A C T G T G T G G T G G C T A C C T G T T A A A T T T G G G A T C G A A C A 100200
K E I C G A H G I P F K E N A D V S N E T F D N Q V S H Q S A V Q K F N P I S C 25

C G G T C C A C A T G C A A T C C C A C C C A T T G A A G C A A T G C C T G G T A C G G A A G G G C A A C C C G A A A C A T T A T G T A C G A A G G G T C G A T T G G A G T T A T A T A A C A C C C A A T C G 100320
- I V S M D 301

59
P G R H L G M G M Q L L A T P V S P P L E S F M 1

A T C T C G G G T C G C T T A C G C T A A A A T A C T A T T G G C T T G A A C G A A A T G C G A A T T C C G A A A T G A A C A C G G G A C A A T G G C A C G G A T G C G C T G T G T A G C A C C A G A T G A C A T T 100440
I E P E G K R T F Y E N A Q V F H R C N R F P V R S L P S P H A H T L V L H C R 261

G A A T T C G G T G G G T T G T C T T G T G C A T G C G A C C C C A C A G C A T A A A A C T A A C C T G T A C G G T T C T C G C A T A A C C T G T A G C A C G G T T G C A C C A C G C C C C C A G C C T A A G T A T A C A 100560
S N P Q T T X Q A H A G M L M F V L G T R N E C L R Q L V R Q V L R G M G L Y V 221

T G C A C C C G G A G T C C C G C A C A C C G T A A G C G T G T A T T C A G A A T A A C A C C C C T G C C T T G C C A C T C C C A G G A T C C G T A G T G G G C G A G T C A T A T T T G G G A T G A T T C C A T G 100680
H S G P T G R R V T L T N L L L V G Q R A W S E L C G H T P P T M N P Y S E M 181

A G G G C G C A A A A T A T T T T A A G A C T A G A C G G T G T A T G C C A C G T T T A C A T A A C G C T A G C C C A T G T G C A T G T C C C G G T A G G G T A T G G A T C T T G A C C A A T A T T A C A C G C G A 100800
L A A F I N K L S S P P T I G R K V S F A L G N A H G A T P Y P D Q G I I V V R 141

A T G C T C T G G G T C C G A A A A T C G C T C C A T G A A A A A T A T C G C T G T A G A T G G A A G T A T T T C C C C T G A A T T A A A G A C G A T T G A T T C T A A A A A A T A C C T T T C G T A C G G C T C T 100920
I S Q P G C F R T W A F I D G T S P L I E E G S M L L R N Y E L F I G K A Y P E 101

T T A A G T C G T C G C A A C A G G T C A T A C C A C T A C G G G A A T G T T A A C T T G T G A A A C T T C A A C C G A A T C C A G T T G C G A A G A C G G G G T G A A C G T T T C C G T G C T A A T G A T G T G A C 101040
K L E D S L L D Y M E P S I N F K S F V E V S D L Q S S V P T F T E T D V H H S 61

A T G T T A T T A C T T G A A G G T G G G G G T A G C T T A A C C C C A A A G G C A G C C G G G G T C G T T G C G G G T T T T T T G G T A A C C G A T G G G C A A A C A A A T G T C C T T T G A T C C G A T 101160
M N N L K F T P P D L K V G L P L G R P R K R T K K T V P H A L V Y I D K S D S 21

A G T T T C A T T C A T T G C A T A C G C G T G G A A C A A C G G T C C C A C A C A C A C C A T T T T C C G G A T A T T T G T G G A A G A T G A G A G T C A C C A T A C A C C G A A A G G C A T C C A A 101280
L K M E N A Y A N S C V T P E G S V D M 1

60
- Q C V R Q F L R D A G M V C G N E P I N T S S P T S D V W V G S L A D L 124

C A A A G C A T C G C G A T G T C C C C G T T T A T G T C T C A C C A C A G A T T G T C C A C C C C T T T A A G G T A C G A T T G G A T T T G C A G T A C G C A T T T G T G C T T T A A C C A A A G T A T A A C 101400
L A D R I D G S K H E E G V S Q A L G K L T V Y P N T W Y A M Q K D K F W L I V 84

T T C G G A C T G G A C A T T T T G C T T A A C C A G A T C C G A T A C G C C T C G T G A G G T T T G A T A C C G G G T G C C G A T A G T C C C A C G C C T A T A C C G A T G A C A C G A C G G T C C G T T A T 101520
E P Y P C K T K V V I G S L A E S L N S V P P A A Y D W A E Y V S S V C P E T I 44

AATCAACTCACATCCGATAGCGGTTGGCTCCAAAAACACGGAGTGTCTTGGAGATGAAGACAATACGGATTGTAGATTGTTAAAAAACTATCTGCAGTAACCATTTATG 101640
I L S V D S L P K A G F L P T D D Q L H L C Y A I T I T K L F V I Q L L W K H 4
TGATGCCATGACGCTTGTTTCCCTTCACTACGAGCTTGTGTATCTTTGAAAACTTGACCACTTAATGGAAGCATGGAAGAATGAGTTTATATATACAGTTGCGCTTTAGT 101760
S A M 1
TAACTCTTGGTGCATATCTCATTTTCTAAAAAGGCGATCTTAATATGTCAACGTCACGGCGTCCGACAAAGCGAATTTCCATGCAAGATTGGATGTAGTATTTATACACCCAA 101880
TCACATGTCACGTATAAGCTTTACAGTCCCCGTTATCTGATATAATCACTTTTCTTAACACGTCATCGGGAACAGATGTTTATATTACCTCTCGCGGTCTTTACGGCAATAC 102000
TTAGACCGTTTCAAGCGGACTGAAAAACGCTCAAAATGCCCTTTTGGAGGCTGCCCAACGGCCATTATCCCTTGGATCTAAGATTGATTGCGGTAACTTTGCCAATCAAGCTTTAAAA 102120
ACGTACCCCAACTTAAACGCTCAAAATGCCCTTTTGGAGGCTGCCCAACGGCCATTATCCCTTGGATCTGAGATTGATTACGGTAACGTTTGCCAACCACGCAATTTACGTTTAAA 102240
TATTTCTAAGCATTCTTAGTGCCTACTTGGCAGCGTGTAAAAATCAACCAATATCCATTATGCTACACGTTTCTTCTATCCGTTTCAATCCATTAAAAAGTCCATTAAACAAAAATGA 102360
TGCAATACCTAATTCACCTAAAAACCTGACTCATTGCAGCAGCGTTTCTCTTGCAGACTATCCAGTTGGCATTTTAAACGGGTCCGGCTGCTAAACCGAAAAACCGTTGCCCTT 102480
ACTGTAAGTCAAAACTAAATTTATATTGCGTGCCTTATTTGTAACATATATGCCCTTTTATCCCCCGAAGTTTGTATCCCTGCCCTTACCACCCCGCCACCTTCCGCCCTT 102600
TTAATAACTTTAATGCTATAAGACATACCCAAACCGGATGTTTTTGGCGTGGAAAAACAGCTTCTAATTTTCCGCTCACTCGGCTTGGTGGATCTCCAAGTATACCTTTAGT 102720
TTGCTCCGTAGAGGTGATAAATACAAACGGTGACAAGTATTGAGCGTAATCTCAATTTTGTATTTAGGGGCGAGCGTTACGACAGCAGATGCGTACTGTTAGAGTGTATGTTT 102840
ATTGTATTGACAGCAGGATGCCCGGTTACTCGAGCAGGATTGCGGCAATCCGAATCGTGTACGGACTTACGAGGGGAGTATTACACCTTGGGTCCAGATATACCAACCTT 102960
TACGACCAATAGCAACACTCAGGTATTTTAAATGCAGCTTAATGATCATAATTTACATACAGTTGGTAATAAGCAGACTGCGATGTTAAGGCAATTCCTTCCCCCTCCCAACAA 103080
ACTAGGACTTCTTATCTTGTGGAATACCTTTACCGCTTTACCGGAGAGCTTTTTTGGTAAGGTGTTTCAGTGAACGTGATGTTGATCCGAGGTGGAGGGGATTGGACTCC 103200
- S K K M K N P I G K G A K G A S S K K T L H K L S G S T S G S T S P T N S E G 429
CCTGTGGAGAGCACTTTCGGGTTTACTTCCCTTACATGCGGAATCAGACTCAGATGTCAGGTCTATTGTTAAGCATCGTTAAGCTCTCTGCGGTATGAAATAAACGGCGCTTAG 103320
Q P S A V K R T K S G K C A S D S E S T L D I T L C R K V D R G T H F L R R K A 389
CACCCCTTCCGCTTCCCGTTTAAATCCCGTAAACAGAAAAAGCTGACTTTTGGGGTGTATTTACCAATCGGATATCCCTTTCATCGCCACGAGAGGTCTCCCGGTTGAGGTG 103440
G R A S G P K I G P L V S F L R V K Q P T N V L R T D R E D G R S T E G T S T 349
TTTCTGGTCTTACAATTGGACCTGTAATAGTTGGATGGCTGTATCTTCCAGGTCCAGGTTGCTAGGTTAGGCGGTTGGATCGGTACATCGATCCCAAGAATAACATGTTGTTA 103560
E P R V I P G T I L Q I A T D K M T W T Q M T L R T P D T C R D L L F L M N T V 309
CAACGGTCTGTTGAATCGAAAAAGACACGAGGATGTTTTAATCCCGCTCATCAGCGCGTAATACCTATATAGTTTAAATACCAATTTTGTAGGCTCTACAATTCG 103680
F P G T S D H L L C R L S T K L G A E D R G Y I G I Y N L I L M K T P E V I E P 269
GTTGATACAGTTCCGCAAGTTGATCATCAAGCATCCGAGTAAGGTTGATGTAACAGGGAATCTCGGCTTCCCTCTGTTCTCTATCCGTTGCTGAAAAGGCAAGTCTGCTCATG 103800
Q Y L E A L Q D D L W G L L P Q M Y C P F R A N G E T G R D T A R F P L R D W T 229
TTGCTGGGCTTGTATTAATCCACAGATCTGACGATCAGGATGCTGCGCCCGGTCGGGGTGTGCTGAGATCAATCGAGCAATACACACCGGGTGGCGGATCGAACAG 103920
R P D Q N I G V S V P R D R Y D Q G G T R P Q Q A S E I S G Y V V P T A S R V A 189
CAGGTTGGCTTTAAAAAATCTTCCGTAATAATGATGCGGTAGAGCATGTTTTGTTACACAGGGCTCGAGTCTCGGGTGGTGGTGTATAGAACTCTGTTGAGAGTCACTTGGT 104040
P Q D K F F V K R L F S A T S C T K T V G P S S D R T P P Q I S D Q Q S D S P S 149
ACTCTGCTGGGCTCTTAGCCGACGATTGAAGGGGCCAGGTTTGGTGAATGGGCTCCGACTCGATCTTGTATGTTGGTGTGGATGGACTCCGACTCGGCTTGGGCTTG 104160
E A T P E R A S S Q L P G P N P S Q I P E R S S R S T P Q Q I S E R S P G P S P 109
GTGCGAAGATCTATGACATCTCCGGTAGGATGTCGATGGAATCTTCAATGACGGCTCAGAAAAACATCGTCTCGGATGGGTGCACTTCATATCTTGTAACTGTATCACTTA 104280
P L L D I V D G P L I D I S D E F S P E S F G D O D S P H V E Y E K Y S T D S V 69
CGATCTTATGAGGATGATTGCACTGGACACCGGAGAGGACACTGGACGCTGGTGGAGGTCCATGCCGAATACAAACAAAGCAAGCTGTCACACAGGCACTGGTTTTCGGA 104400
I K H L I S Q V P C R C L P C Q V S T S T W A R I C V F C F D H L C P M T K G L 29
GATCGAAACGGTCTCATGATATGGTGCAGGTATTATCCGAAGCTCGAGGTGCCGTACCGCCGCTAATATGGTATCCATGTAACAACTGGCTGTATCTAATGTCGGGCATC 104520
D S V T S M C I T C T N D S A D S T G S G G A L I T D M 1
CAACACGTAGCAGAACTGCCATGCGTTCTAAATGTGAGTTGTGGCGAGTACATTTTATAATGGTACCAACGAGACACACCCCTATATCCCTCCACCATTTCTTTAAGTCCAC 104640
CCACTAAACGTGGGTATAAATGTGATTGGGTAGGCGGACAGTCCCAACAAACAGGGAAGTTGATTGTTAATACCTTGGCCCGGATACAGTAAGTGACATTTAGATTCTGTCT 104760
TTATTAGATAAAGAGCATACGAAGACATTTCCACCCCTGTAATACCGTAATAAAGGTAAGTCCCAACAAAGACACTGTATATAGGAAGTCGGGTGATTGGGACAGTTAC 104880
UL ----- IRL
TCCATTAGAGCGGTACAAACAACTAGGATAGGGTAATGCAAGTCCCCCGATGGTGCCTCCGCAACGCGGGGAGGTGGGTCGCTTTTTTCTCTCTCGAGGGGGCCGCGG 105000

IRL ----- IRS
GAGGGTGGCTTCTCCCGGGTCCCGGGGCGCCAGAAACCGGGGGGGTTATTTTCGGGGGGGGTCCGACCGCCCGCTCGCCCGCCGACAGACAGACACTTTT 105120
TTCAAAAACCGTTCGCTTTTATTAACAACAAACAGTCCGCGCCAGTGGCGCTACAGAAAAAGGAGGGGACTCCGTCACCCCGACTTCGCGGGGGCTCTCCCCCGGCGCT 105240
- G R S Q P P S R G G R G 1299
CCCCACACATGCTCTGCTCGGAGGACGAGGACGAGGACAAACAGTCCACTTGACCGCGGGCGAAACCCACCGGGCTCTCGAGCACACCGGGGCCACCGACAGATGTC 105360
G M V D D E D E S S S S S L L E V K V A P R L G V R R D R L V G P A V S V I S 1259
ACCCAAAGGATGACCGGGTGGTCCCCGCTGCTCCCGCGCCCTCTCGCTGCTCCACGCGTTCACACCCACCTCCCAATGCTCCAGCTCCAAGGCTGTTCTCTGCTGCTGG 105480
Y G F S S G P A D G D D G G G E E S D W A D E C G V E W D D L E L A H E R D D A 1219
GTGCGCGCTGTCGCGCCGCTGGGTTTGTACGCGCTTCGAGCGCCGCTGGTGTCGGAACAGAACCGTTCCTGCTGCTCCCTCAACACCGTCTCCCGGGGCCAAACCGGGC 105600
T R R Q R G A Q T E S P R E S G G H H G F V F R T G D S G E L V T E A A G F G P 1179
GGCCACATTACTGGAATCGGGGAGGGGATTTCGAGCTGCTCGCGCCGATACAGCGCCACCGACCGCGCCACGGGTGGAAGCAGAGTGGTCTGCGCGAGGGTGGG 105720
P N M V R P I P P L A N R A E D A S A Y L A V S R G A V P P L V L P E A A P D P 1139
TCCAGCGGGGCTGCGGCAAAACACCTCGCCAGGTGGGTACGTCGCGCGCTCCGGCCGGCGCCCGCGCTCCGCTCTCGGGAAGGAAGCGGGTGAAGCGCGGACCGG 105840
E L L A H R C F V R A M T P V D G A E P G A A G P R R G E P L F V P R L A A G L 1099
CCCCATCGGTTGTGCGGGTGGTATGTGCGCTGCTCCCAAGTGGGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 105960
G M R N A A R H S H A A E D V F D A A G L G L G R S Q R A L D K C G D F G P L V 1059
TACTCGCGTATTACGGGGCGACAGGGGACGCGGGTTCGGGGCGCGCGGGTACACAGGTGTATGCGACTTCCACCGCGGCAAAACAGGGGTTGTGCGCGGGTACAG 106080
Y Q R Y E R P S L P V R T K P G A R T C V T V A V N G G R C L C L P Q E G P Y L 1019
TTGCAAAACGAGTCTGATACGAGCAAACTCGTGGCCAAAGGTGCGGACGATGCAACACGCGCGGGGAGTCTTCTGTGACCGCGAGTCTGGCCATCGGACGAGCGCTGG 106200
N A F A T E I R A F S A P G F T R S S A F V A R A L G E T V A S D P M R V V A Q 979
GCGTCGCGCGCGGGCGCGGACGATACGCTGATGAGACAAAGCGGGTCCATCCCTGGGCGCCCTCTCGAGGGCCACCGCGTCAACACCAAGAACCGGGCGGCGGCGGCGG 106320
A D P R A P A R V V Y H Y Q S L A P G D R P W R E L A V A D L V L L R R R A S A 939
AACCGGAGCTAGATCTGACGCGCCCGGCAAGGCGAGGTCTCGGGTGCAGTAATAAAACGCGCGGCGTCAAGCGGACAGCTCCGCGGGCGGTCAGTCTCCGGCCCA 106440
L R S G L Y E V A G A F A L D R T S L L L V G R A N L A S V D P P G T W N G A W 899
GCATGAGTGTGCGGAGGACAAACGGTACTCAGGGTGCAGGACACAGAGTCCCTCGGGATGACTCCATGACGGTCCCGGATCTGTGCGAGGGTGTCTCGAGGGGGCG 106560
A H T S P L C L R N S L A A L V S L G G R S P S W S P G P D T A L T S E L P G 859
TTGATGCTCTCGCGCAACGATCTGATGATGAGAGCCCTACATCTCCGCGTGGGATCTGCGCATCCAGGCGACCTCCGTCGAGCGCTCCACTCCGCTGGGTGAGCA 106680
N I D E G P L P D Y I I L L R V D E P D P I Q R M W A C R R R L A E V G S P P G 819
AACGTCGGTCTCTCGCGCGGCGGCGGAGCGGATTCGCGCAAGGCGCGGATCAAGCTTAGCGAGGCGCGAGGCGTGGGAACAATGGTGTGCGACGAGCGGCGATG 106800
F R R D G G G P R R A A I E A L A G P D F S L A P R W A T P F L P D D V L R A I 779
GTTTCGGGGTACAGTACGCTTGCAGCTGCTGCGAGCGGATGCGAGGCGGCGGATGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 106920
T E P T C Y A K R A Q D S P V P T H L A G R P I R R F G G N P A P G D L P M T I 739
ACGGCGGGGGATCCACACAGGGCGGAGGTGATGCTGCGGCTGCGATACCGGCTCTTGGCTTGAACACCATGATGCTGCAACCGGGCGTCCGCGAGGGTGTCTCCCTA 107040
V A P P D V V P G S T I T V P E S V R R K A K S V Y H D D A V R A D A V P T E R 699
ATCTGTGAGGAGGCTTCTGCTCTGACTGCTGGGACTTGCCTTGCCTGCGGAGTCTGTAACGATCATCGGTGACACAGAGAGAGCGTGGCGGCGGCGGCGGCGGCGG 107160
I K D L L S R S E V P Q S K R K R P T R L R D D P P C V S L S R A A A S P Q P D 659
GGAGCTGTGTGCGGGGTTGTGGAGAGGGTACCGGGGAGTCCGCGCGCGGAGTGGAGCGCGTTCCTCGGGGTATGCAATGCTGCGCAAGGCTCTGCGGAGACTGTGAGG 107280
P A Q T A P T T P S P H G R S I R A A P S S G T A E P Y A M S A F A R R L S Q L 619
ATAAGTGTGTTGGGCGGCTGATGACGCGCTATAGCCACGCGCGGCGGCTGGGGGAGAGCCAGAGGCGCTCCCGGTGCGCATGCTTCCCTACATGCGGAACGGGAGAC 107400
I F H K Q A R D Y R R S M A V A A A H P L A W L A E G T A M A E G V H P V P S 579
GCTACGCTCCCGTAACGGCGTACCGCGCTGCTGAGCGGATCCCACTCCGCGAGTAAAGCGGACGGGTTCGTTGAGGCTCTGAGGCTATGCTATAGGCTCT 107520
A V S G T V A T G A R G P P L L K Q Y F Q N L A S N V G T L K P N Q L M A I P D 539
CTGTCTGAGAGTAGTACGTTAATCAGCGCGGCTACTGTAGCGGATCTCCCACTCCGCGAGTAAAGCGGACGGGTTCGTTGAGGCTCTGAGGCTATGCTATAGGCTCT 107640
R D P C Y I L N I L A R Y Q R A P D G L E P V Y L P V P E T S A E Y R A R A A R 499
ACAGCTCATCTCCAGTACGCTCTCTGCTGCTCCCGGAGGTCAAAACCGACCTGTTGGATGGAGGGGTGCGGATCCGCGGCAAGGGCTCCGTCGGGATCATGAGCGGCGG 107760
V A E D E M H G E R T E G S P G F R V R N S P L P A S G P W P S G D P M M L P G 459
GACACCGGGGAATTATCGGGTCTGATCGCGGAGGAAATGATTCTGTCTGCGCGCGGTTCCCGCGAAGAGCTTTGGTCTTACGAATCTCGGATCGGAGCGCTGATG 107880
S V P P I I P T R S R P L S F S K Q R Q R G P E G A L R K T K R I R P D P G S I 419
GATCGATATCCGGTGGATATTTTGTTCGAGCCACCATCTTGTAGTCCGAATCATCGAATTTGACGGGAGGGGCTGTTGCGCTCGGAGCTGCTGCTGATGTTCACTT 108000
S R Y G P Q I N Q K T S G G D N S D S D D S N S P S P A H E R G S R S G T T E S 379

[illegible]

1799

W T P S P K R P Q R A I E R Y A G A E T A E Y T A A K A L T A L G E G G V D M K	260
AACACCCCGTCCCCAAAGACACCCGACGCTGCATCGAGCGATACGCGGGTGCAGAAACCGCGGAATACACGCCGCGAAAGCGCTACCGCGTTGGCGAGGGGGGTGTAGATTGGAAG	11350
RRRHEAPRRHDIPPPHGV -	278
CGAGCTGCACAGAGCCCCCGCGCGCATGATATACGCCCCCCCATGGCGTGTAGTCTTTATAAATAACAATGGTTGGCTGCTGCTTTTTTGTATGCTGCTGCTGCGGGAGT	11480
M N L C G S R G E H P G	12
GGGGTGTGTGATATTAGAGGGTAGAGGGTGTGTTTGAAGCTCTCCATTAAACCCAGGGGTCGCCACCGGGCGTGTGGTATGAATCTCTCGCGATCCGCGGTGAGCACCCGGG	11600
G E Y A G L Y C T R ' H D T P A H Q A L M N D A E R Y F A A A L C A I S T E A Y E	52
GGTGAATATGCGGACTTTATCGACACGACAGATACCCCGCGCACAGGCTCTCATGAACGACGCCGAACGGTACTTCGCGCGCGCTATGCGCATATCTACCGAGGGCTACGAG	11720
A F I H S P S E R P C A S L W G R A K D A F G R M C G E L A A D R Q R P P S V P	92
GCTTTATACAGCCCTCCGAGAGACCGTGCAGGATTTGTGGGGAGGGCAAGGACGCTTCGGACGATGTGCGGGAGCTCGACGGATAGACAACGTCCACCCTCGGTCCG	11840
P I R R A G L S L L R E Q C M P D P Q S H L E L S E R L I L M A Y W C C L G H A	132
CCGATCCGAGAGCGGTGTATCGTATTATCGCGAGCAATGCATCGGATCCACAATCGCATCTGGAGCTCAGCGAGCGGTGATATTGATGGCATATTGGTGTGTTTGGGACACGCC	11960
G L P T I G L S P D N K C I R A E L Y D R P G G I C H R L F D A Y L G C G S L G	172
GGACTTCGCACTATTGGATTGTGCCCGATAAATGCATCCGCGCGAATTATATGACCGCCCCGGGGAAATTGTACAGGCTTTTGAACGCGTACCTGGGCTGCGGGTCCCTTGA	12080
V P R T Y E R S -	180
GTCCCAAGAACCTACGAGAGATCTCGACACCCCATCCCTTTATAGAAAAAATAAATTTAAACATACACCGGATAAAGCGTACTGTTTTTATTTAAATTCACGCTCGCG	12200
TTGCCCGGTTGCGTGATCAGCGGCTTTATCTATATACACCGTGTAACGAAACCCCGTGACTCCCTCCAATCGGTTACCAAACTCTTCTCGGTATCGGTAGATTCCGAGTCTCG	12320
JRS ----><----US	
AAATCGTCCCACTTATCCCAAAATTTGTGACGTTATATATCCCAAGGCAAGGCGCTCCCGTCATAGCAAAATACAAGACAATTTAGCGTAATATAACAGAATTTTTTACGATGATAT	12440
- G V F Q S T I Y G L A F A A G T M A F V F V I I L T I Y C F K R H Y	68
ATTTTATGTTGATATTTTCCAATTGACGCAAAAAATTCATCTGCCGTTTCATTTTCGCTATCACTATAAACAACCTTTTCAGCGGAACGGCTGGTGTATGCGTGTATGCTGTTATTA	12560
I K H Q Y K G I R R L F E D A T E N E S D S Y V Y C K E A S R S P Q I A T I T N	28
TTTGGTTCGCTCGCGGGGTACCAACGCTTCCATCAGTAAGGCCACGGCTCACCCCTCCATGGTGTGTTTTCGCGGCATAGAAATCCAGATTGTAAGGCCAGCAGGCTAGTTTAAAGT	12680
M P Q A R P T V V A E M L L A V A E G E M T N Q G A M	1
GTTTAATACCACACCTTTTGATATTATATACATGCAAGATTCTAGATTATTCATCAATAGGTCGTTTAAAGCGGTTTCATAACGTTGTCAGCTATACCGACATTCACAAAGAGG	12800
TAAAGTTACCTTACGTTATTATTAATAAATAACATGTAGACATTATTAATAATCTAGGAACAATCAATCCATATTGTAAGTATTGTTAAACCCCTCCCTTTTGTGCATTATCTCCGC	12920
M	1
CCTCTTATAATCGGATCACTTTATAAGTGCTGCGTGAGTATATTTGTACAGTTGTGGACACAGGTTTTGGTTCATTAACTATCAACATAAGTCGGGGTATACAAGTATAATGA	13040
N D V D A T D T F V G Q G K F R G A I S T S P S H I M Q T C G F I Q Q M F P V E	41
ACGAGCTGTGATGCAACAGACACCTTTGTGGACAAGGAAGTTCCGTGGCGCATCTCAACATACCGCTACATATTATGCAAAATGTTGGGTTTATACAACAGATGTTTCCAGTTGAAA	11360
M S P G I E S E D D P N Y D V N M D I Q S F N I F D G V H E T E A E A S V A L C	81
TGTCGCCCGCATAGAAATCGAGGATGTCCCAATTATGACGTTAACTGGATATACAGCTCTTTAATATATTTGAGGTTGACACGAACTGAAGCCGAAGCCTCTGTGGCATGTGCG	113280
A E A R V G I N K A G F V I L K T F T P G A E G F A F A C M D S K T C E H V V I	121
CAGAAGACCGGTTGGAATTAATAAGCGGGATTGTAAATATAAAGCAGTTTACACAGGGCGGAAGGTTTGGCTTTGGCTGTATGGACAGTAAACATGTGAACATGTGGTCATTA	113400
K A G Q R Q G T A T E A T V L R A L T H P S V V Q L K G T F T Y N K M T C L I L	161
AGCGGGCTCAACGTCAAGGAACCGCCACGAGGCAACCGTGTAAAGAGCTTAAACCCACCATCCGTTGACAGCTTAAAGGAACGTTACGTATAACAAATGACATGTCTTATATTAC	113520
P R Y R T A D G L Y L A A K R N L P I C D I L A I Q R S V L R A L Q Y L H N N S	201
CAGCTTACCGAACAGTATTATACGTATCTAGCTGCAAGGCGCAACCTCCCCATATGTGACATTTTAGCAATTCAGGCTGTGATTACGCGGTTACAGATCTTCTATAATAACAGTA	113640
I I H R D I K S E N I F I N H P G D V C V G D F G A A C F P V D I N A N R Y Y G	241
TTATTCACCGTATATAAATCTGAAAATATATTATTAACCAACCGAGGTGATGTTTGTGGGAGACTTTGGAGCAGCGTGTTCCTCCCGTGGATTAATAGTCCAACAGGATATTATGGCT	113760
W A G T I A T N S P E L L A R D P Y G P A V D I W S A G I V L F E M A T G Q N S	281
GGGCTGGAACAATCGCCACAACCTCTCTGAGTTATTGGCTAGAGATCCATATGGAGCTGCCGTGGACATATGAGTGGCGGGATTGATATTGAAATGGCTACAGGACAGAACTCGT	113880
L F E R D G L D G N C D S E R Q I K L I E R S G T H P N E F P I N P T S M L R	321
TATTTGAAGAGAGCGGTTAGATGGCAATTGTGACAGTAGCGTCAAAATAAATCTATTATACGACATCTGGAACCTACCCAATGAATTTCCCATTAACCTTACATCAAACTCTGCTC	114000
R Q Y I G L A K R S S R K P G S R P L W T N L Y E L P I D L E Y L I C K M L S F	361
GACAATACATTTGGTTGGCAAAACGGCTTCTCGAAAACCGGATCCAGGCCATTGTGGACAATCTATATGAGTTGCCAATTGATTGGAGATTGTGATATGTAAGATGTTATCGTTTG	114120
D A R N R P S A E V L L N H S V F Q T L P D P Y P N P M E V G D -	393
ACGACGCTATCGACCATCAGCAGAGGTTGCTTAACCACTGTGTTTCCAACTCTCCGATCCATATCAAAATCCATGGAAGTTGGAGATTAAATTCATTAAAGCTGTTAATAA	114240

VZV DNA sequence

1801

T T K P K E I T P V N P G T S P L L R Y A A M T G G L A A V V L L C L V I F L I 558
 G A C T A C T A A C C C A A G G A A T T A C C C C G T A A C C C C G A A C G T C A C C A C T T C A C G A T A T G C C G A T G G A C C G G A G G C T T G C A G C A G T A C T T T T A T G T C G T A A T T T T T A T 117480
 C T A K R M R V K A Y R V D K S P Y N Q S M Y Y A G L P V D D F E D S E S T O T 598
 C T G T A C G G C T A A C G A A T G A G G T T A A A G C T A T A G G T A G A C A G T C C C G T A T A C C A A A G C A T G T A T T A C G T G G C C T T C C A G T G G A C G A T T T C G A G G A C T C G A A T C T A C G G A T A C 117600
 US ----- TRS
 E E E F G N A I G G S H G G S S Y T V Y I D K T R - 623
 G G A A G A A G A G T T T G G T A A C G C A T T G A G G G A G T C A C G G G T T C G A G T T A C A C G G T G A T A T A G A T A A G A C C C G G T G A T C A C C G A A C C G G G G C A A C C C G A C G C G T A A T T T A A T A A 117720
 A A A C A G T A C G C T T T T A C C G G T G T A T G T T T A A T T A T T T T T T C T A T A T A A A G G A T G G G G T G T C A G G A T C T C G T A G G T T C T T G G G A C T C C A A G G G A C C C G A C C C A G G T A 117840
 - S R E Y T R P V G L S G C G L Y 165
 C G C G T C A A A A G C T G T G A C A A T T C C C C G G G C G G T C A T A A T T C G G C G G G A T G C A T T T A T T C G G G C G A C A A T C C A A G T C G G A A G T C C G G C G T G T C C C A A C A C C A A T A 117960
 A D F L R H C I G G P R D Y L E A R I C K N D P S L G I T P L G A H G L C C W Y 125
 T G C C A T C A A T A T C A C G C C T C G T G A G T C C A G A T G C A T T G G A T C C G G A T G C T G C G G A T A A A C G A T A A C C C G T C G C G G A T C G G C G A A C C G A G G G T G A C G T T G T C T 118080
 A M L I L R E S L E L H S Q P D P M C Q E R L L S L V A R R I P P V S P P R Q R 85
 A T C C G T C G G A G T C C C C G C A C A T C C G T C G A A G G C T C C T T G C C C C C C A A A C T C G G C A C G G T C T C T G G A G G G C T G T A T A A A G C C T C G T A G C C T C G G T A G A T A T G C 118200
 D A A L E G C M R G F A D K A R G W L S A C P R E S P S H I F A E Y A E T S I A 45
 G C A T A G C G G C G G C G A A G T A C C G T T C G G C G T C G T A T G A G A G C T G G T G C G G G G G A T C G T G T G T G C A G T A A G T C C G G C A T T T C A C C G C C G G G T G C T A C C G C G G G A T C C 118320
 C L A A A F Y R E A D N M L A Q H A P T D H R T C Y L G A Y E G G P H E G R S G 5
 G C A G A T T C A T A C C A C G G C C C G T G T G G G A C C C C G T G G G T A A T G G A G A C G T T C A A A C C A G C A C C C T C T A C C C T A A T T C C A A C A C C C C A C T C C C C A G A C A G A C A T A A A 118440
 C L N M 1
 A A A G A C A C G A C C A A A C C A T T G T A T T T A T A A A G A C T A C A C C C A T G G G G G G C G T A T A T A T G C C G C G G G G C T T C G T G C A C G T C G C T T C C A A T C A C A C C C C T C G C C 118560
 - V G H P P P I D H R R P A E H R R R K M D V G G E G 253
 C A A C G G G T A G C G C T T T C G C G G T G T A T A T T C C G G T T T C G A C C C G C G T A T C G G T G A T G C A C G T G G G G T C T T T T G G G A C G G G G T G T T G C A C C A T C C C G T C G A T A A C A 118680
 L A T L A K A A T Y E A T E A G A Y R E I A R Q P R K P S P T N C G D G D R Y C 213
 A T C G C G G G T A C C C G T C T G G T T C A A A G A T C G G T G C T C C T C T G A T T C T T C C A C T A T A A A G T C T C C C G T C C G A T T C C G C T G G A A T C C G A A T T C A A T A C A T C C G A 118800
 D G P V G D P E C S D T S E E S E E E V I F D E G D S S E A D S D R F E I V D S 173
 T G C G T A G A C C A T C A T C G C T A T C T T C C A C C A T C A T C A G A T A C G T C A C A T C C A G T G C C T C T A T G C A A G G A A C A T T C G G C C C A A T G A A C C G T C T T C C C C G A G G 118920
 P T S G D D D S D D E G G D D S V D C E L A S R H L P V N P A E I F G D E G R P 133
 G G T C G A T G G G A C A C T T C C A T T A A G A G C G G T A G C A C T G C G T G G G T G A A G C T G A A C A A T T C C T C C A G C A C G T A A C G T G G G A C G T G T T G G T G C G A C C C A T A G A T A 119040
 P R H S V E M L T A R L C Q T Q T P Q L Q L L E E W C A L T P R Q R T R G M L Y 93
 A A A G T C A G G C A T A T G C C G C A A T T C T G A A T C G C G T A C G C G T C C C G A C G A C T G C A A T A C A A A A A A C A C G G T G A T C C T C A A C A G C C T G G C C G G G T G A T C C C A 119160
 F D L C I R R L R R S D R T Y R G S S Q A I C L F V R N I D E V L R A P H I E M 53
 A C C C G A G T A C A A A A C G C C G G G C C C G A C G A T A C C G G C G T T C A G G G C C T G G T G C A G A T C C A T T C C A T T T C C G T T A A C A T C A A C C G A T G C C C G G G T T T G A C T C G G A 119280
 G P T C F A G P G R S T D R G N L P G P A S G Y E M K G N V D V S A G P K S E S 13
 C G A G T C G C C C C G C G A T C C G G T G A G G T G C A A A A C A T G T C T T G G G C C G T A G T A A C C T T T C C C T T A A A A C C A C T C G A C G T G T C G C G T T A T G A A T C G G A C G A C C C T G C A A C A A A A 119400
 S D G R T A P S T C F M 1
 C A C A C C C A A C G T T A C A T A T A A A G C T A C T G G G T A A A T G C A A T G G G G A T T C C G G G C G G A G A C C T T C G A T T G G G T T C C T T T A T A A C C A A A A A A A G G G G G G G C C 119520

 C C G T G T T T T T T T T A C A G T A A T C A A T T T A A A A G C C T G C C G T C C A T T T G G A A T A T A T A T T C T G T G A A A G C C C C C A C A C C C A T A A A A C C G C A C A T C G C G G G A A C A C 119640

 -----origin of DNA replication-----
 G C G G A A C A A G A A C T C T C T C T T T C T A T A T A T A T A T A T A T A T A T A T A T A T A G A A A G A A G T G C A A C G T G G T G G A C A C A T G C C A A A C A T G A A A C C C A T A C A G T A A 119760

 A A A C G G A A G T G C A A T G C A G A T A A A A G A G T A T C C G A T T G G C G T A C A C C A G A C A T G C G G A C G C C A A T T T A A C C C C C C C T T T T T A C C C C C C A C C C A C C C A T T C A C C C 119880

 C A G G A A G T G C G A A C G G G T T A C A T G C C T C A G A T A T G A A G T T C T C G A C T G T T T T T G A A T A A T T T T T T G T G A T T T T C A A C G G T T A G A G A T T A T G T T A T A A C A T C G C G G G 120000

 -----reiteration R4-----
 < A' X A' X A' X A' X A' X >
 T A C C G C G C C C C C C C A T C G G C G G G T A C C G C C C C C C C C C C A T C G G C G G G T A C C G C C C C C C C C C C A T C G G C G G G T A C C G C C C C C 120120

 A >
 T C C C A T C G G C G G G G T A C G T A A C A C C A A C C C G T G T A T T T A T G G T T A T C G G G T C T G T G C C C T G A C A T A A T C G T T G G A G G G G T G G T G T A T A C G T T G T G A T 120240
 T G C G C A A C T A A T A C A G C A G A G A G G G C C C A A C A C C C G T G A C G T G C A T T T G A T A A C T A G A T G C C G A T G G G T G A A A C A C C C G T T A T A A G A T G T T T T G C A T G T A G A C A 120360

ACCCAAATTGTGTTATATATATCGTCTGAGACACAGATGTTGGTGTTATTAACAATATGTAATAAATGCATGTCGTGGTATCCCCGTGTATGATGTTGAAGT 120480

ATGCGAAATAGACACCGGGCGTACATCGCAACAGCGGTCCTCCTTAAACGCATACTATGTCCTCAAGAATCCCGGCTCGAGTCGTCCAATCACTACATCGTCTTATCATTAAG 120600

AATATTTACACGGTGACGACACGGGAGGAATATCGGTCGAGGGGGGGCACACAGCTTTTAAGTAGTCTTGGAACTCCCTACCAACGCCAATCGCAATCCCTTTGAAGGCTCGGAG 120720

M D T P P M O R S T P Q R A G S P D T L E L M D L L 26

AGCGTTTGAAAACCTGGGATCGTCTAAATTCACCCAGTCGCGATGCGATACGCGCGCATGACGCGCTACACCCCAACGCGGGGTCGCTGATACITTTGGAGTTAATGGACGTGTT 120840

D A A A A A A E H R A R V T S S Q P D D L F G E N G V M Y G R E H E I V S I 66

GGACGCGGCGCGCGCGCGCGCGAACACAGGCGCGGCGGTCACCTCGAGTCAGCTGACGATCTACTATTTGGAGAGACGGGTCATGCTGGGACGGGAAACAGAGATGCTTCAAT 120960

P S V S G L Q P E P R T E D V G E E L T Q D D Y V C E D G Q D L N G S P V I P L 106

TCCCCTCGTATCGGAGCTCAACCAGAACCCAGAACGGAAGATGTTGGCAGAGCTAACACAAGAGCTACGATGCGAGGAGCGTCAGGATCTAATGGGCTCGCGTGTATCCCGCT 121080

A E V F H T R F S E A G A R E P T G A D R S L E T V S L G T K L A R S P K P P M 146

GCGCAGGCTCTCCACACCGGATCTCGAGGCGCGCGCGGAGAACCAACAGAGCGCGATCGCTCCCTTGAGACGTCTCTCTCGGAACGAACTGCTAGGCTCCAAAAACCCGAT 121200

N D G E T G R G T T P P F P Q A F S P V S P A S P V G D A A G N D Q R E D Q R S 186

GACGATGGGGAACGGGACGAGTACGACCCCTCGGTECCGCGAGGCTTCTCCCTGTATCCCCCGCTCTCTGTTGGAGACGCGCGGGAACGATCAACGGGAAGACGCGGTC 121320

I P R Q T T R G N S P G L P S V V H R D R Q T Q S I S G K K P G D E Q A G H A H 226

TATACCCGCAACAGCAGCAGGAAATTCACAGGTTTGCCGTCGGTGGTCCATCGAGACAGACAACTAGTCCATCTCGGTTAAAAAGCGGGGATGAGCAAGCGGGTCATCGCA 121440

A S G D G V L Q K T Q R P A O G K S P K K T L K V K V P L P A R K P G G P V 266

TGATCGGGGACGAGTAGTCTCCAGAAAACCTAACGCGCGCTCAGGGAAGAGCCCGAAGAAAAGACTTTGAAGGTTAAGGTCCTCCCGCGCGGAAACCGCGTGGACCTGT 121560

P G P V E Q L Y H V L S D S V P A K G A K A D L P F E T D D T R P R K H D A R G 306

ACCCGCGCGGTTGAGCAATTGTACAGCTCTTCTCGACAGCGTTCGCGTAAGGGGGAAGCGGAGCTGCGGTTTGAGACCGATGATACCGCGCAAGGAACATCATGCGGGG 121680

I T P R V P G R S S G G K P R A F L A L P G R S H A P D P I E D D S P V E K K P 346

TATAACACCTCGCGTCCCTGGACGTCTGCTCGGGGGCAAACTAGAGGTTTTTGCCCTGCCGGAAGATCCACGACAGACCGGATGAGGATGACAGCCAGTGGAGAAAAGCC 121800

K S R E F V S S S S S S S W G S S S E D E D D E P R R V S V G S E T T G S R S 386

AAAGAGTCGTGAGTTGTTGCTCTTCACTCTCTCTCGTCGTGGGATCGTCAAGAGATGAAGACGATGAACCCGCGCGGTTGCGTGGGAAGTGAACATACAGGACGAGGTC 121920

G R E H A P S P S N S D D S D S N D G G S T K Q N I Q P G Y R S I S G P D P R I 426

CGGACGCAACAGCGCCCTTCCCGCTCAAATCGGATGATCGGACTCAAATGATGGTGGTCGACGAAACAAATATCCAACCGGATATGATCCATCAGCGGTCGCGATCCGAGGAT 122040

R K T K R L A G E P G R Q R Q K S F S L P R S R T P I I P P V S G P L M M P D G 466

TCGTAAGCAACAGCTCTGCGGGGAACCGGGGCGCAGAGACAGAAATCATTTTCCCTGCCGCGATCCGAACCCGATAATTCGCCGCTGCGGGGCGCGTATGATGCCCGACG 122160

S P W P G S A P L P S N R V R F G P S G E T R E G H W E D E A V R A R A R Y E 506

AAGCCCTTGCGCGGATCGGACCCCTCCATCAACAGGTCGGGTTGGACGCTCGGGGAGACAGAGAGGCTACGAGGATGAGGCTGAGAGCGCGCGGCTCGTTACGA 122280

A S T E P V P L Y V P E L G D P A R Q Y R A L I N L I Y C P D R D P I A W L Q N 546

GGCTCAACGGAACCGCTGCGCGTTACGTCGCGAGTGGGAGATCGGCTAGACGATACCGCGCGGTGATTAACCTGATCTACTGTCCAGACAGAGACCTATAGCATGGCTCCAGAA 122400

P K L T G V N S A L N Q F Y Q K L L P P G R A G T A Y T G S V A S P V H V G E 586

CCCCAAGCTGACCGGTTCAACTCGGCGTGAACAGTGTACCAAAAGCTGTTGCCACCGGACGGCGGATACCGGCTTACGGGAGCGTAGCTCTCCGTTCCGATGTAGCGCA 122520

- A M A T G E A L W A L P H A A A A V A M S R R Y D R A Q K H F I L O S L R R A F 626

AGCCATGGCAGCGGGGAGGCTCTGGGCTTCCCCACGCGCGCGCGGCTGGGTATGAGCCGTCGATACGACCGGGGCAAAACACTTTATCTCAAGATGCTCCGAGAGGCTT 122640

A S M A Y P E A T G S S P A A R I S R G H S P T T P A T Q A P D P Q S A A A 666

TGCGAGATGGATACCCGAGGCAACGGGCTCAGTCCGGCGCGGAGATCCCGGCTACCTTCTCAACACCCCGGACACAGGCTCCGACCTCAGCGCTCGGCGCGCG 122760

R S L S V C P P D D R L R T P R K R K S Q P V E S R S L L D K I R E T P V A D A 706

ACGCTCTCTTCTGTGTGTCACCGGATGATGTTACGAATCCGCGAAGCGCAAGTCCGAGCTGAGAGGACGAAGCTCTCGACAGATAGGAGACACCGCTCGCGGACG 122880

R V A D D H V V S K A K R R V S E P V T I T S G P V V D P P A V I T M P L D G P 746

CGGGTTGACAGCATGATGTGGTTTCAAGGCCAAGAGCGGGATTCGAGCGCGGACCATACCTCGGCGCTGTGGTGGATCCCCCGCGTAATAACGATGCCACTTGACGGACC 123000

A P N G G F R R I P R G A L H T P V P S D Q A R K A Y C T P E T I A R L V D O P 786

GGCCCCAAACGGGGGATTTGCGGATTTCCCGGGGGGCTTGCATACCCGCTCCGTCGCGACGAGCTGCAAGGCGTACTGTACCCCGGAACCATGCGCGCTGTGGTCGACGACC 123120

L F P T A N R P A L S F D P G A L A E I A A R R P G G G D R R F G P P S G V E A 826

ATTGTTTCCACGGCTTGGCGCTTGGCTAAGCTTTGATCCGCGCTTGGCGGAAATCGCCCTCGGCTCGGCGGAGAGACCGAGGTTGGTCCACACGAGCGGATGAGGCG 123240

L R R R C A W M R O I P D E D V R L L I Y D P L P G E D I N G P L E S T L A 866

GCTGCGAGGAGTGGCTGTGATGCGGAGATCCGAGCCGAGGATGTGAGGCTTCTGATCATCTACGATCGGTTGCGCGGAGAGGACATCAACGGCCCCCTGAGAGACCTCTCG 123360

T D P G P S W S P S R G G L S V V L A A L S N R L C L P S T H A W A G N W T G P 906

GACAGATCCGGACCGTATGAGGTCATCCGAGGGGACGTGCTGTGGCTTGGACGCTGAGTAACCGGTTGCGCTGCGGACGATCATGCTGGGCGGGAACGTGACCGGGCC 123480

P D V S A L N A R G V L L L S T R D L A F A G A V E Y L G S R L A S A R R R L L 946
 G C C G G A C G T G C C G C T T T G A A C G C C G G G C G T T T A T A C T G C A C C G A G A C C T G G C C T T T G C C G G G C C G T C G A G T A T C T A G G C T C G C G G T T G C C T C T G C C C G G C C G G T T G C T 123600
 V L D A V A L E R W P R D G P A L S O Y H V Y V R A P A R P D A Q A V V R W P D 986
 G G T G T G G A C G G G T G C C C T C G A G A G G T G C C C A G G G A T G A C C C G C T T T G T C A G T A T C A C G T A C G T C C G G G C C C G G C G C A C C G A C G C C C A G G C C G T C C G A T G C C A G A 123720
 S A V T E G L A R A V F A S S R T F G P A S F A R I E T A F A N L Y P G E Q P L 1026
 C T C G G C G T C A C A G A G A C T C G C C G G G C G T G T T G A T C G T C G C G A C C T T T G G C C A G C A G T T T G C T G T A T C G A G A C T G C G T T T G C C A A C T G T A C C C G G G C A A C A C C C T 123840
 C L C R G G N V A Y T V C T R A G P K T R V P L S P R E Y R Q Y V L P G F D G C 1066
 G T G T T G T C C C G G T G G A C T C G C A T A C C G T G T A C C C G C G G G C C C A A G A C C C G C G T C C C C T G T C G C C C C G T A A T A C C G C A G T A C G T G C C G G G T T T G A C G G T T G 123960
 K D L A R Q S R G L G L G A A D F V D E A A H S H R A A N R W G L G A A L R P V 1106
 C A G G A C C T C G C G A C G T C T C G G G T C T G G G C T C G G G C A C C G A C T T T G T G A C A G A G G C G C A T A G C C A C C G C A G A A A C C G A T G G G C C T G G T G C C G C G T T C G A C C C G T 124080
 F L P E G R R P G A A G P E A G D V P T W A R V F C R H A L L E P D P A A E P L 1146
 C T T C C T C C C A G G A G G A G A C C G G G G C C C G G G C G G A G C C G G C A C C T G G C G A G G G T T T T T G C C C A C G C C T G T G A A C C G A C C C T G C C G C A G A C C A C T 124200
 V L P P V A G R S V A L Y A S A D E A R N A L P P I P R V M W P P G F G A A E T 1186
 C G T G T C C A C C C G T G C G G T C G T C G T G G C G T A T G C G T C G G C G A G G G C T C G A A T G C C C C C C G A T T C C A G A T A A T G T G C C C C G G T T T T G G G C C G G A G A C 124320
 V L E G S D G T R F V F G H H G G S E R P S E T Q A G R Q R R T A D D R E H A L 1226
 G T G T T G G A G G G A G C G A A C A C G G T C G T G T T C G G A C C A C C G G G C T C G A A C C C A G G C G G G C G A C A G C G G C A C C G C A G A C A G A G A C A C C G T T T 124440
 E L D D W E V G C E D A W D S E E G G G D D G D A P G S S F G V S I V S V A P G 1266
 G G A G C T G A C A T T G G G A G T G G G T G T A G A C C G C G T G G A C A G C G A G G A G G G G C G G G A C A C G G G A C G A C C G G G T A T C C T T T G G G T G A G A C T G T G T C G T G G C C C C G G G 124560
 V L R D R R V G L R P A V K V E L L S S S S S E D E D D V W G G R G G R S P P 1306
 T G T G C T G C A G A C C G C G G G T G G T T G C C C C G C G G T A A G T G G A G C T G T T G T C C T C G T C C T C C G A G G A C A G A G A C A T G T G T G G G A G G G C G C G G G G A G G A C C C C C 124680
 Q S R G - 1310
 G C A G A G T C G G G G T G A C G G A G T C C C C C T C T T T C T C G T G A G C C A C T G G C G C G C G A C T G T T T G T T A A A A G C G G A C G G T T T T A T G A A A A A G T G T C T G T C T G T G C G G 124800
 G C G G C G A C G G G C G G C T G T C G A C C C C C C C G A A A A A A C C C C C C C G T T T C T G G G C G C C C G G C G A C C C C G G A G G 124884

Fig. 1. Sequences of the VZV genome and encoded proteins. The DNA sequence is shown as the rightward 5' to 3' strand only. The leftward strand would have an additional C residue at the 3' end and lack a C residue complementary to the G residue at the 3' end of the rightward strand (Davison, 1984). The IR_L-IR_S junction would also be displaced one residue to the left on the leftward strand. Rightward encoded protein sequences are shown in single letter amino acid code above the corresponding DNA sequence. Leftward encoded proteins are shown below the DNA sequence. ORFs are designated by number at the left of the first line containing the amino acid sequence, regardless of coding orientation. The last nucleotide of each AATAAA-related element predicted to function in transcript polyadenylation is indicated by an asterisk above or below the DNA sequence. The locations of the ends of VZV dPyK mRNA are indicated between genes 35 and 37. Four potential TATA signals and three AATAAA-related elements potentially involved in polyadenylation are underlined. The two ATG codons in the untranslated 5' region of this mRNA are marked with asterisks; they are in different frames from the initiation codon for dPyK and are followed by termination codons. A copy of the VZV DNA sequence will be deposited in the EMBL sequence library.

identical in two independent overlapping clones (*KpnI* c and *HindIII* a). R1 is not located in a recognized region of size variability between virus isolates (Straus *et al.*, 1983).

It is possible that regions other than those containing R2, R3 and R4 are variable in size, as the analyses of VZV isolates would not have identified regions where size differences are small or very infrequent in occurrence. For example, McGeoch *et al.* (1985) described a tract of G:C base pairs in the U_S component of HSV-1 which varies in single base pair steps. Nonetheless, the results described above imply that the VZV genome is not unique in size. The extent of genome size variability among different virus isolates, and the DNA sequences of the reiterations, indicate that the genome may vary in length from just above 124 000 bp to more than 126 000 bp.

The discovery of tandem G + C-rich reiterations in a herpesvirus genome was first made with HSV-1 (Davison & Wilkie, 1981; Watson *et al.*, 1981b), and the role of these sequences in genome size variation was first established with this virus (Davison & Wilkie, 1981). However, it is clear from the HSV-1 sequence data already available that VZV has far fewer reiterations

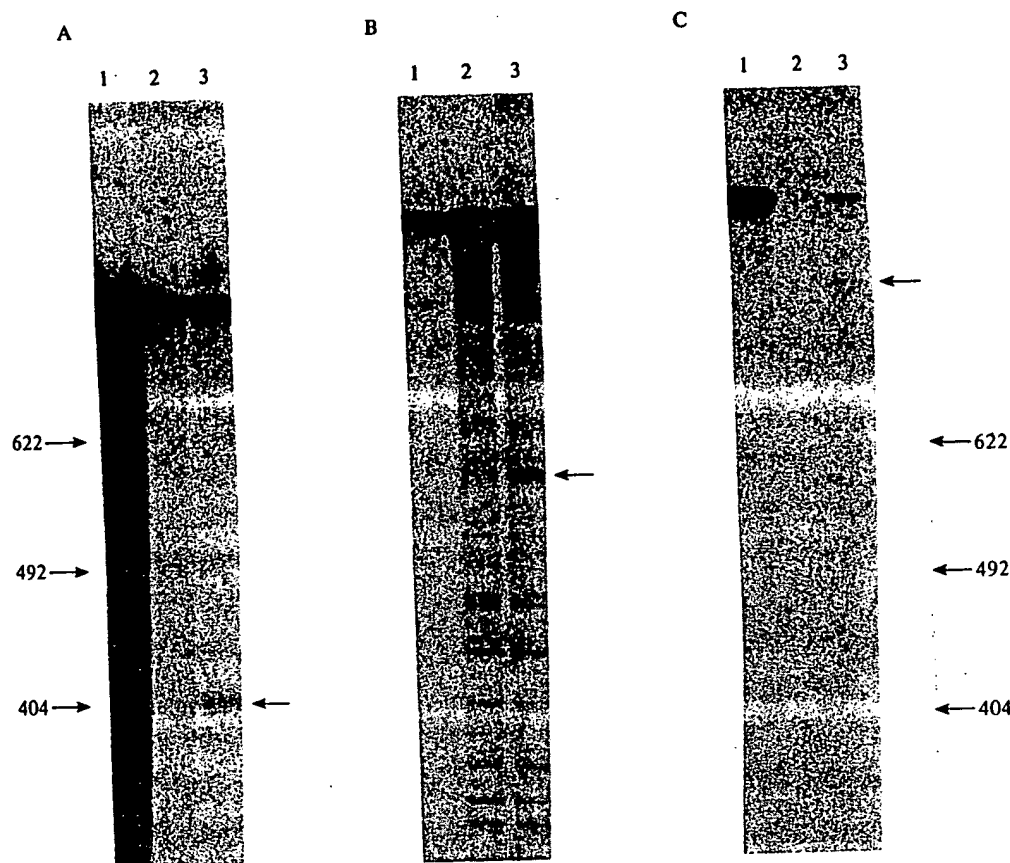


Fig. 4. S1 nuclease analysis of VZV dPyK mRNA. The three end-labelled probes used were (A) 881 nucleotide fragment 5'-labelled at the *AccI* site and extending leftwards to a *PstI* site, to detect the 5' end of dPyK mRNA; (B) 1275 nucleotide fragment 3'-labelled at the *XmaI* site and extending rightwards to a *PstI* site, to detect the 3' end of dPyK mRNA; (C) 1390 nucleotide fragment 5'-labelled at the *XmaI* site and extending leftwards to a *PstI* site, to detect the 5' end of dPyK mRNA. Each set of three lanes shows (1) untreated probe; (2) S1 nuclease treatment of probe which had been incubated with uninfected cell RNA; (3) S1 nuclease treatment of probe which had been incubated with VZV-infected cell mRNA. Lane 1 contained one-tenth the amount of probe represented in lanes 2 and 3. The sizes of DNA markers in nucleotides are shown on each side of the figure. DNA fragments protected by VZV-infected cell RNA are indicated by arrows.

proposed for some genes whereas no good candidates may be found for others. Thus, the confident identification of a potential promoter for every VZV gene is precluded.

Gene 36 provides a salutary lesson in this respect. The product of this gene has significant homology to HSV-1 deoxypyrimidine kinase (dPyK), and is therefore an excellent candidate for the VZV dPyK detected experimentally (Doberson *et al.*, 1976). The molecular weight of the predicted protein (37815) is in good agreement with the experimentally determined value of 35000 for the monomer subunit (Lopetegui *et al.*, 1983). Fig. 4 shows the results of mapping the ends of the dPyK mRNA, and their locations with respect to the DNA sequence in Fig. 1 are indicated in the region of gene 36. The S1 nuclease results indicate that the 5' end of the mRNA maps 410 bp upstream from the *AccI* site (Fig. 4, lane A3) and 920 bp upstream from the *XmaI* site (Fig. 4, lane C3), consistent with the function of the sequence TATATAA underlined in Fig. 1 (64364) as the TATA element. Three similar sequences (TATATTA, TATAAAA and TATAATA) in Fig. 1 (64433, 64469 and 64611) are present between this element and the initiation codon for dPyK. Thus, the location of the 5' end of dPyK mRNA could not have been

predicted from the DNA sequence alone. Similarly, the 3' end of the mRNA might have been expected to map near the AATAAA or ATTAAA elements underlined in Fig. 1 (66125, 65978). In fact, the 3' end is located 580 bp downstream from the *Xma*I site (Fig. 4, lane B3), and is thus defined by the AGTAAA element (65859). This result counsels caution in predicting the 3' ends of other VZV mRNAs; nonetheless, the general degree of confidence was sufficient for proposed elements involved in polyadenylation to be included in Fig. 1 and 2. In any case, there are precedents for the function of AGTAAA in polyadenylation (Donehower *et al.*, 1981; Tamura *et al.*, 1981; Capon *et al.*, 1983). The possibility that the dPyK AGTAAA resulted from a base change in an AATAAA during cloning was ruled out by sequencing this region in an independent clone (*Sst*I *h*). The element AGTAAA has also been proposed in Fig. 1 and 2 to function in polyadenylation of the transcript from gene 28, which encodes the DNA polymerase.

The dPyK gene is interesting in other respects. Although it encodes a protein with substantial homology to HSV-1 dPyK, as described below, the untranslated 5' region of the mRNA is considerably larger, at 420 bp, than that of the HSV-1 mRNA, at 110 bp (McKnight, 1980). Moreover, the untranslated 5' region of the VZV gene contains two ATG codons in different reading frames from the initiation codon for dPyK. Each of these is followed within a short distance by a termination codon and, in view of the work of Kozak (1984), this structural aspect may affect translational efficiency of the mRNA by requiring reinitiation for expression of dPyK. As the only VZV gene promoter identified experimentally to date is that for the dPyK gene, responsible comparisons between HSV-1 and VZV promoter regions may be made only for this gene. Little similarity was detected between regions upstream of the TATA elements. The differences between the structures of the promoters and untranslated 5' mRNA regions of the VZV and HSV-1 dPyK genes suggest that transcriptional and translational control might differ considerably between the two genes.

The locations of potential polyadenylation sites near the 3' ends of ORFs are summarized in Fig. 2. Many genes apparently possess unique polyadenylation sites, whereas others are present in 3'-coterminal families containing up to four genes. The 3'-coterminal gene arrangement has been well-characterized in HSV-1, where sets of genes are expressed as mRNAs with unique 5' ends but shared 3' ends (for review, see Wagner, 1985). Thus, the mRNAs expressed from genes towards the 5' end of a 3'-coterminal family contain extensive untranslated 3' regions. The VZV genome contains 216 AATAAA elements, but only 48 unique potential polyadenylation sites are predicted in Fig. 1 and 2. Moreover, seven of these sites contain ATTAAA rather than AATAAA, and two contain AGTAAA. It is possible that the presence of some AATAAA-related elements close to the ends of ORFs is merely fortuitous, and thus some genes shown with unique polyadenylation sites in Fig. 1 and 2 actually may be members of 3'-coterminal families. This comment applies particularly to VZV mRNAs predicted to have unique 3' ends whose HSV-1 counterparts are polyadenylated as part of a 3'-coterminal family. For example, HSV-1 transcripts corresponding to VZV ORFs 40 and 41 are 3'-coterminal (Costa *et al.*, 1981), as are those corresponding to ORFs 43 and 44 (Costa *et al.*, 1984). Ostrove *et al.* (1985) recently reported a transcript map for the VZV genome, based on Northern blot analysis using relatively large cloned DNA fragments. The arrangement of VZV mRNAs suggested by this approach correlates well with that deduced from the DNA sequence in some regions, but there are apparent discrepancies in others. Confirmation of the transcript map and resolution of these differences must await the mapping of mRNA termini.

Although identified overlapping polypeptide-coding regions are not extensive, the VZV genome shows considerable economy in its gene layout. Almost the entire sequence encodes virus proteins, and it is likely that many regions involved in control of gene expression are located in the coding regions of adjacent genes. However, there are four notable regions for which no protein products are predicted. One is located at the left end of the L segment and contains TR_L. It may contain sequences which promote cleavage of full-length genomes from concatemers produced during DNA replication (Davison, 1984). The second region, also about 600 bp in size, is located at the right end of the L segment and contains IR_L. Part of its function is likely to be as a promoter for gene 61. The third region is about 1400 bp in size and is located

between genes 60 and 61. It contains an unusual direct repeat of 88 bp, with three mismatches, separated by 24 bp (102020–102219 in Fig. 1; A indicates each repeat). The function of this structure is unknown; it could form part of a control element for gene 60. The fourth region is about 1400 bp in size and is present twice in the genome: in IR_S between genes 62 and 63 and in TR_S between genes 70 and 71. It contains the reiteration R4 and the promoters for the two genes on either side. It also contains a palindrome (Davison & Scott, 1985) which forms part of a functional origin of DNA replication (Stow & Davison, 1986). In view of the compact arrangement of the rest of the genome, these four regions are likely to have important functions. Although no protein products have been assigned to them, it is possible that some contain small coding exons or perhaps larger non-coding exons. Alternatively, they may encode functional RNAs which are not expressed as proteins. A third alternative is that they may encode no RNA or protein species, but are sites for specific recognition during the virus life-cycle. This is certainly the case for the origin of DNA replication, and probably for sequences at the left end of the genome potentially involved in DNA maturation.

VZV gene function

Comparison of the proposed arrangement of VZV genes in Fig. 2 with published HSV-1 transcript mapping data (for review, see Wagner, 1985) indicates that both viruses have a similar gene layout. This view was confirmed by available HSV-1 sequence data, and allowed the functions of several VZV genes to be assigned on the basis of primary amino acid sequence homology of their products to HSV-1 proteins. These conclusions, and the precise locations of VZV genes and molecular weights of their primary translation products, are summarized in Table 1. Genes encoding glycoproteins, homologues of HSV-1 immediate-early proteins, and proteins with extreme properties of hydrophobicity, hydrophilicity, charge or amino acid composition are also indicated. All but three of the functional assignments were made on the basis of HSV-1 gene location and confirmed by amino acid sequence homology with HSV-1 proteins. The dUTPase was assigned on the basis of the location of the HSV-1 gene reported by Preston & Fisher (1984). The thymidylate synthetase and protein kinase genes were located on the basis of amino acid sequence homology of their products to proteins of known function in the NBRF protein database. Approximately 30 VZV proteins are homologous to proteins predicted from the complete EBV sequence determined by Baer *et al.* (1984); the implications of this result in predicting the functions of EBV genes will be discussed elsewhere (A. J. Davison & P. Taylor, unpublished data).

Fig. 5 shows examples of homology between VZV and HSV proteins displayed by optimal alignment of predicted amino acid sequences. Fig. 5(a) shows a comparison of the product of VZV gene 18 with the small subunit of the HSV-2 ribonucleotide reductase; these proteins are highly conserved. The lower degree of homology between the VZV and HSV-1 dPyKs shown in Fig. 5(b) is in accord with the DNA hybridization data of Davison & Wilkie (1983), who were able to detect conservation of the ribonucleotide reductase gene but not of the dPyK gene. The degree of homology shown in Fig. 5(c) between the product of VZV gene 5 and the potential HSV-1 membrane protein reported by Debroy *et al.* (1985) is about the same as that observed for the dPyKs. However, several pairs of genes are less conserved than this, and only specific regions of the proteins were detected as being conserved by this approach. Fig. 5(d) shows the conservation of a region towards the carboxy termini of the glycoprotein product of VZV gene 14 and HSV-1 glycoprotein C (gC). Although the homology in this region is significant, the major part of each protein is divergent. The divergent region of the VZV protein contains a repeated amino acid sequence coded by reiteration R2.

Most glycoprotein genes encode primary translation products with distinct characteristics: a hydrophobic signal sequence near the amino terminus for translation of the mRNA on membrane-bound ribosomes (Blobel, 1980) and a more extensive hydrophobic region followed by basic residues near the carboxy terminus for anchoring the protein in the membrane (Tomita & Marchesi, 1975). The VZV genome contains five such genes: 14, 31, 37, 67 and 68. Genes 14, 31, 37 and 68 are counterparts of identified HSV-1 glycoprotein genes, as shown in Table 1. Gene 67 is the counterpart of an HSV-1 gene whose predicted product also has the

Table 1. *Properties of proteins coded by predicted VZV genes*

Gene	Start*	Stop†	Residues	Mol. wt.‡	Extreme properties§	Function
1	915	592	108	12103	Hydrophobic (C)	
2	1134	1847	238	25983		
3	2447	1911	179	19149		
4	4141	2786	452	51540	Hydrophilic (N)	Homologue of HSV-1 IE63 ¹
5	5274	4255	340	38575	Hydrophobic	
6	8577	5329	1083	122541		
7	8607	9383	259	28245		
8	10667	9480	396	44816		dUTPase
9	11009	11914	302	32845	Hydrophilic	
10	12160	13389	410	46573		trans-inducing factor ²
11	13590	16046	819	91825	Hydrophilic & acidic (N)	
12	16214	18196	661	74269		
13	18441	19343	301	34531		Thymidylate synthetase ³
14	21113	19434	560	61350		Glycoprotein (gpV); homologue of HSV-1 gC ⁴
15	22478	21261	406	44522	Hydrophobic	
16	23794	22571	408	46087		
17	24149	25513	455	51365		
18	26493	25576	306	35395	Acidic	Small subunit of ribonucleotide reductase ⁵
19	28845	26521	775	86823		Large subunit of ribonucleotide reductase ⁶
20	30475	29027	483	53969		
21	30759	33872	1038	115774		
22	34083	42371	2763	306325		
23	43138	42434	235	24416	Hydrophilic; S, T, Q-rich	
24	44021	43215	269	30451	Hydrophobic (C)	
25	44618	44151	156	17460	Hydrophilic; acidic (N)	
26	44506	46260	585	65692		
27	46127	47125	333	38234	Hydrophilic & basic (N)	
28	50636	47055	1194	134041		DNA polymerase ⁷
29	50857	54468	1204	132133		Major DNA-binding protein ⁷
30	54651	56960	770	86968		
31	57008	59611	868	98062		Glycoprotein (gpII); homologue of HSV-1 gB ⁸
32	59766	60194	143	15980	Hydrophilic & acidic	
33	62138	60324	605	66043		
34	63910	62174	579	65182		
35	64753	63980	258	28973	Basic	
36	64807	65829	341	37815		Deoxypyrimidine kinase ⁹
37	66074	68596	841	93646		Glycoprotein (gpIII?); homologue of HSV-1 gH ¹⁰
38	70293	68671	541	60395		
39	70633	71352	240	27078	Hydrophobic	
40	71540	75727	1396	154971		Major capsid protein ¹¹
41	75847	76794	316	34387		
42	78038	76854	395	82752		
43	82593	81538	352	(spliced)		
44	80360	81448	363	40243		
45	82719	83315	199	22544	Hydrophilic & acidic	
46	83168	84697	510	54347		
47	84667	86319	551	61268		Exonuclease ¹²
48	86226	86468	81	8907	Hydrophilic	
49	87882	86578	435	48669	Hydrophobic	
50	87881	90385	835	94370		
51	90493	92805	771	86343		
52	93850	92858	331	37417		
53	95984	93678	769	86776		
54	95996	98638	881	98844		
55	98568	99299	244	27166	S, T-rich	
56	99626	99414	71	8079	Hydrophilic & basic	
57	100272	99610	221	25093	Hydrophilic & basic	
58	101219	100305	305	34375		
59	101649	101173	159	17616	Acidic	
60	104485	103085	467	50913	Hydrophilic	
61	109133	105204	1310	139989		Homologue of HSV-1 IE175 ¹³
62	110581	111414	278	30494	Hydrophilic & acidic	Homologue of HSV-1 IE68 ¹⁴
63	111565	112104	180	19868		
64	112640	112335	102	11436	Hydrophobic (C)	
65	113037	114215	393	43677		Protein kinase ¹⁵

characteristics of a glycoprotein (McGeoch *et al.*, 1985). The glycoprotein products of VZV genes 31, 67 and 68 have been identified unequivocally: they encode gpII (Keller *et al.*, 1986), gpIV (Davison *et al.*, 1985) and gpI (Ellis *et al.*, 1985), respectively, according to the recently established VZV glycoprotein nomenclature described by Davison *et al.* (1986). Thus, there are two probable VZV glycoprotein genes (14 and 37) whose products have not yet been identified, and one antigenically defined major glycoprotein (gpIII) whose gene has not yet been mapped. In the absence of additional data, the product of gene 37 has been tentatively assigned as gpIII? in Table 1, and that of gene 14 has been proposed as gpV, a previously undetected minor glycoprotein.

The VZV genome also contains four genes (5, 15, 39, 50) which encode particularly hydrophobic proteins. Mutations in the HSV-1 counterpart of gene 5 result in a syncytial plaque morphology, and it has been suggested that the product of this gene is a membrane protein (Debroy *et al.*, 1985). Thus, it is possible that VZV gene 5, and perhaps genes 15, 39 and 50, encode membrane-associated proteins.

Although the HSV-1 and VZV gene arrangements are similar, there are limited regions of significant difference. The most extensive encompasses the S segment, which in HSV-1 contains 13 unique genes (McGeoch *et al.*, 1985, 1986a) and in VZV contains only seven. The relationship between the S segments of VZV and HSV-1 has been discussed separately (Davison & McGeoch, 1986). In summary, each VZV gene has a homologue in HSV-1, but the remaining six HSV-1 genes have no counterparts in VZV. The 'missing' genes include the immediate-early gene which encodes HSV-1 IE12, and the gene which encodes glycoprotein D. These regions of the two genomes differ substantially in gene layout, but they are clearly related, and a scheme has been proposed for their descent from the S segment of an ancestral herpesvirus by expansion and contraction of the inverted repeats. Two other regions of difference between the VZV and HSV-1 genomes are present at the ends of the L segment. These regions have been sequenced in HSV-1 (L. J. Perry & D. J. McGeoch, personal communication), and appear not to contain homologues to VZV gene 1 and perhaps gene 2, at the left end, and gene 61 at the right end. Also, the inverted repeats flanking U_L are much larger in HSV-1, at approximately 9000 bp, than they are in VZV (88.5 bp), and the single gene thus far identified from the HSV-1 sequence of this region specifies a spliced immediate-early mRNA encoding IE110 (L. J. Perry, F. J. Rixon & D. J. McGeoch, personal communication). At the present stage of analysis, no homologue of this gene has yet been detected in VZV. Thus, the differences in gene arrangement between VZV and HSV-1 in the S segment and at the ends of the L segment result in VZV apparently lacking homologues to two of the five HSV-1 immediate-early genes: those encoding IE12 and IE110. There is only one other region for which sufficient HSV-1 data are available to indicate a difference in gene layout between VZV and HSV-1. It is almost certain, from several lines of evidence, that HSV-1 lacks a homologue to VZV gene 13, which encodes a protein with a

67	114496	115558	354	39362		Glycoprotein (gpIV); homologue of HSV-1 US7 ¹⁴
68	105808	117676	623	69953		Glycoprotein (gpI); homologue of HSV-1 gE ¹⁴
69	118332	117793	180	19868		
70	119316	118483	278	30494	Hydrophilic & acidic	Homologue of HSV-1 IE68 ¹⁴
71	120764	124693	1310	139989		Homologue of HSV-1 IE175 ¹³

* Location in rightward strand of first base of initiating ATG codon. All except 14, 31 and 68 refer to the first ATG in the ORF (see text).

† Location in rightward strand of last base of the codon preceding termination codon.

‡ All predicted mol. wt. values except those of the primary translation products of genes 14, 31 and 68 (see text), were calculated from the amino acid sequence commencing at the first in-frame ATG in the ORF.

§ (N) and (C) indicate that the extreme property is limited to a region towards the amino or carboxy terminus, respectively.

|| References are given in the superscripts to HSV sequence data, or other sequence data for genes 13 and 66, which confirm the assignment of VZV gene function. ¹L. J. Perry & D. J. McGeoch, personal communication. ²Dalrymple *et al.* (1985). ³A. J. Davison & R. W. Honess, unpublished data. ⁴Draper *et al.* (1984). ⁵Y. Nikas & J. B. Clements, personal communication. ⁶McLauchlan & Clements (1983). ⁷Quinn & McGeoch (1985). ⁸Bzik *et al.* (1984). ⁹McKnight (1980). ¹⁰U. Gompels & A. C. Minson, personal communication; McGeoch & Davison (1986b). ¹¹A. J. Davison & J. E. Scott, unpublished data. ¹²Draper *et al.* (1984); McGeoch *et al.* (1986b). ¹³McGeoch *et al.* (1986a). ¹⁴McGeoch *et al.* (1985). ¹⁵McGeoch & Davison (1986a).



Fig. 5. Optimal alignment displays of the predicted (single-letter) amino acid sequences of (a) the product of VZV gene 18 and the small subunit of HSV-2 ribonucleotide reductase (McLauchlan & Clements, 1983); (b) the product of VZV gene 36 and HSV-1 strain 17 dPyK (D. J. McGeoch, personal communication); (c) the product of VZV gene 5 and the potential HSV-1 membrane protein described by Debroy *et al.* (1985); (d) the product of VZV gene 14 and HSV-1 gC (Draper *et al.*, 1984). In each example the VZV sequence is shown above the HSV sequence.

remarkable degree of homology to prokaryotic and eukaryotic thymidylate synthetases (A. J. Davison & R. W. Honess, unpublished data). Instead, Frink *et al.* (1983) have shown that this region of the HSV-1 genome contains a small gene which is present in a 3'-coterminal family with the gC gene (the counterpart of VZV gene 14) and is thus in the opposite relative orientation from VZV gene 13. Thus, although VZV and HSV-1 are very similar in gene layout in the L segment, this discovery enhances the possibility that one or more other local differences may exist.

The VZV and HSV-1 genomes also differ in another functional aspect. The region between the HSV-1 DNA polymerase and major DNA-binding protein genes contains a large palindrome (Gray & Kaerner, 1984; Weller *et al.*, 1985; Quinn & McGeoch, 1985) which forms part of an origin of DNA replication (Weller *et al.*, 1985). This origin is termed *ori_L* to distinguish it from *ori_S* and *IR_S*. Plasmids containing the corresponding region of the VZV genome do not contain a palindrome and do not function as origins (Stow & Davison, 1986). Comparison of cloned and virion DNA fragments (data not shown) has ruled out the

possibility that a similar palindrome might have been deleted during cloning, as occurs in clones containing HSV-1 *ori_L*. Therefore, although VZV has an origin corresponding to HSV-1 *ori_S* (110087 to 110350 and 119547 to 119810 in Fig. 1; Stow & Davison, 1986), it does not possess a second functional origin in a location equivalent to that of HSV-1 *ori_L*. However, it cannot be ruled out that VZV has a second origin elsewhere in the genome which may have been deleted on cloning.

Significance of the VZV sequence

The DNA sequence provides a firm foundation on which to build a detailed understanding of VZV infection at the molecular level. This knowledge may be applied in the development of effective treatments for the diseases caused by this virus. The sequence has also given the first complete view of gene layout in a member of the *Alphaherpesvirinae*, and has allowed our knowledge of the proposed functions of VZV genes to increase from almost nothing to equal that of HSV-1. Clearly, the sequence will be important in determining the functions of the majority of VZV genes whose role in virus growth is at present unknown. The way in which data from one herpesvirus may be so usefully applied to another thus encourages herpesvirologists to cultivate a more catholic approach towards the family of viruses they study.

The authors are grateful to John Subak-Sharpe for continuing interest in this work, to Duncan McGeoch for advice and discussion, to Philip Taylor for extensive help with computer programming and to Catherine Joyce for supplying the *E. coli* strain which overproduces the Klenow fragment of DNA polymerase I. Thanks are due also to Duncan McGeoch, Frazer Rixon, Lise Perry, Nigel Stow, Mike Dalrymple, Yannis Nikas, Ursula Gompels and Ed Wagner for making available unpublished HSV-1 data, and to John Subak-Sharpe and Duncan McGeoch for critically reading the manuscript.

REFERENCES

- BAER, R., BANKIER, A. T., BIGGIN, M. D., DEININGER, P. L., FARRELL, P. J., GIBSON, T. J., HATFULL, G., HUDSON, G. S., SATCHWELL, S. C., SEGUIN, C., TUFFNELL, P. S. & BARRELL, B. G. (1984). DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature, London* **310**, 207-211.
- BERK, A. J. & SHARP, P. A. (1978). Spliced early mRNAs of simian virus 40. *Proceedings of the National Academy of Sciences, U.S.A.* **75**, 1274-1278.
- BIGGIN, M. D., GIBSON, T. J. & HONG, G. F. (1983). Buffer gradient gels and ³⁵S label as an aid to rapid DNA sequence determination. *Proceedings of the National Academy of Sciences, U.S.A.* **80**, 3963-3965.
- BLOBEL, G. (1980). Intracellular protein topogenesis. *Proceedings of the National Academy of Sciences, U.S.A.* **77**, 1496-1500.
- BLUMENTHAL, R. M., RICE, P. J. & ROBERTS, R. J. (1982). Computer programs for nucleic acid sequence manipulation. *Nucleic Acids Research* **10**, 91-101.
- BZIK, D. J., FOX, B. A., DELUCA, N. A. & PERSON, S. (1984). Nucleotide sequence specifying the glycoprotein gene, gB, of herpes simplex virus type 1. *Virology* **133**, 301-314.
- CAPON, D. J., CHEN, E. Y., LEVINSON, A. D., SEEBURG, P. H. & GOEDDEL, D. V. (1983). Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. *Nature, London* **302**, 33-37.
- CASEY, T. A., RUYECHAN, W. T., FLORA, M. N., REINHOLD, W., STRAUS, S. E. & HAY, J. (1985). Fine mapping and sequencing of a variable segment in the inverted repeat of varicella-zoster virus DNA. *Journal of Virology* **54**, 639-642.
- CHOU, J. & ROIZMAN, B. (1985). Isomerization of herpes simplex virus 1 genome: identification of the *cis*-acting and recombination sites within the domain of the *a* sequence. *Cell* **41**, 803-811.
- CORDEN, J., WASYLYK, B., BUCHWALDER, A., SASSONE-CORSI, P., KEDINGER, C. & CHAMBON, P. (1980). Promoter sequences of eukaryotic protein-coding genes. *Science* **209**, 1406-1414.
- COSTA, R. H., DEVI, B. G., ANDERSON, K. P., GAYLORD, B. H. & WAGNER, E. K. (1981). Characterization of a major late herpes simplex virus type 1 mRNA. *Journal of Virology* **38**, 483-496.
- COSTA, R. H., DRAPER, K. G., KELLY, T. J. & WAGNER, E. K. (1984). An unusual spliced herpes simplex virus type 1 transcript with sequence homology to Epstein-Barr virus DNA. *Journal of Virology* **54**, 317-328.
- DALRYMPLE, M. A., MCGEOCH, D. J., DAVISON, A. J. & PRESTON, C. M. (1985). DNA sequence of the herpes simplex virus type 1 gene whose product is responsible for transcriptional activation of immediate early promoters. *Nucleic Acids Research* **13**, 7865-7879.
- DAVISON, A. J. (1983). DNA sequence of the U_S component of the varicella-zoster virus genome. *EMBO Journal* **2**, 2203-2209.
- DAVISON, A. J. (1984). Structure of the genome termini of varicella-zoster virus. *Journal of General Virology* **65**, 1969-1977.

- DAVISON, A. J. & MCGEOCH, D. J. (1986). Evolutionary comparisons of the S segments in the genomes of herpes simplex virus type 1 and varicella-zoster virus. *Journal of General Virology* 67, 597-611.
- DAVISON, A. J. & SCOTT, J. E. (1983). Molecular cloning of the varicella-zoster virus genome and derivation of six restriction endonuclease maps. *Journal of General Virology* 64, 1811-1814.
- DAVISON, A. J. & SCOTT, J. E. (1985). DNA sequence of the major inverted repeat in the varicella-zoster virus genome. *Journal of General Virology* 66, 207-220.
- DAVISON, A. J. & WILKIE, N. M. (1981). Nucleotide sequences of the joint between the L and S segments of herpes simplex virus types 1 and 2. *Journal of General Virology* 55, 315-331.
- DAVISON, A. J. & WILKIE, N. M. (1983). Location and orientation of homologous sequences in the genomes of five herpesviruses. *Journal of General Virology* 64, 1927-1942.
- DAVISON, A. J., WATERS, D. J. & EDSON, C. M. (1985). Identification of the products of a varicella-zoster virus glycoprotein gene. *Journal of General Virology* 66, 2237-2242.
- DAVISON, A. J., EDSON, C. M., ELLIS, R. W., FORGHANI, B., GILDEN, D., GROSE, C., KELLER, P. M., VAFAI, A., WROBLEWSKA, Z. & YAMANISHI, K. (1986). New common nomenclature for glycoprotein genes of varicella-zoster virus and their glycosylated products. *Journal of Virology* 57, 1195-1197.
- DEBROY, C., PEDERSON, N. & PERSON, S. (1985). Nucleotide sequence of a herpes simplex virus type 1 gene that causes cell fusion. *Virology* 145, 36-48.
- DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* 12, 387-395.
- DOBERSON, M. J., JERKOFKY, M. & GREER, S. (1976). Enzymatic basis for the selective inhibition of varicella-zoster virus by 5-halogenated analogs of deoxycytidine. *Journal of Virology* 20, 478-486.
- DONEHOWER, L. A., HUANG, L. & HAGER, G. L. (1981). Regulatory and coding potential of the mouse mammary tumor virus long terminal redundancy. *Journal of Virology* 37, 226-238.
- DRAPER, K. G., COSTA, R. H., LEE, G. T.-Y., SPEAR, P. G. & WAGNER, E. K. (1984). Molecular basis of the glycoprotein-C-negative phenotype of herpes simplex virus type 1 macroplaque strain. *Journal of Virology* 51, 578-585.
- DUMAS, A. M., GEELLEN, J. L. M. C., MARIS, W. & VAN DER NOORDAA, J. (1980). Infectivity and molecular weight of varicella-zoster virus DNA. *Journal of General Virology* 47, 233-235.
- DUMAS, A. M., GEELLEN, J. L. M. C., WESTSTRATE, M. W., WERTHEIM, P. & VAN DER NOORDAA, J. (1981). *Xba*I, *Pst*I and *Bgl*II restriction endonuclease maps of the two orientations of the varicella-zoster virus genome. *Journal of Virology* 39, 390-400.
- ECKER, J. R. & HYMAN, R. W. (1982). Varicella-zoster virus DNA exists as two isomers. *Proceedings of the National Academy of Sciences, U.S.A.* 79, 156-160.
- ELLIS, R. W., KELLER, P. M., LOWE, R. S. & ZIVIN, R. A. (1985). Use of a bacterial expression vector to map the varicella-zoster virus major glycoprotein gene, gC. *Journal of Virology* 53, 81-88.
- FITZGERALD, M. & SHENK, T. (1981). The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 24, 251-260.
- FRINK, R. J., EISENBERG, R., COHEN, G. & WAGNER, E. K. (1983). Detailed analysis of the portion of the herpes simplex virus type 1 genome encoding glycoprotein C. *Journal of Virology* 45, 634-647.
- GAROFF, H. & ANSORGE, W. (1981). Improvements of DNA sequencing gels. *Analytical Biochemistry* 115, 450-457.
- GILDEN, D. H., SHTRAM, Y., FRIEDMANN, A., WELLISH, M., DEVLIN, M., FRASER, N. & BECKER, Y. (1982). The internal organization of the varicella-zoster virus genome. *Journal of General Virology* 60, 371-374.
- GRAY, C. P. & KAERNER, H. C. (1984). Sequence of the putative origin of replication in the U_L region of herpes simplex virus type 1 ANG DNA. *Journal of General Virology* 65, 2109-2119.
- HANAHAN, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* 166, 557-580.
- JOYCE, C. M. & GRINDLEY, N. D. F. (1983). Construction of a plasmid that overproduces the large proteolytic fragment (Klenow fragment) of DNA polymerase I of *Escherichia coli*. *Proceedings of the National Academy of Sciences, U.S.A.* 80, 1830-1834.
- KELLER, P. M., DAVISON, A. J., LOWE, R. S., BENNETT, C. D. & ELLIS, R. W. (1986). Identification and structure of the gene encoding gpII, a major glycoprotein of varicella-zoster virus. *Virology* 152, 181-191.
- KINCHINGTON, P. R., REINHOLD, W. C., CASEY, T. A., STRAUS, S. E., HAY, J. & RUYECHAN, W. T. (1985). Inversion and circularization of the varicella-zoster virus genome. *Journal of Virology* 56, 194-200.
- KOZAK, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Research* 12, 857-872.
- KUMAR, A. & LINDBERG, U. (1972). Characterization of messenger ribonucleoprotein and messenger RNA from KB cells. *Proceedings of the National Academy of Sciences, U.S.A.* 69, 681-685.
- KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* 157, 105-132.
- LOPETEGUI, P., MATSUNAGA, Y., OKUNO, T., OGINO, T. & YAMANISHI, K. (1983). Expression of varicella-zoster virus-related antigens in biochemically transformed cells. *Journal of General Virology* 64, 1181-1186.
- LUDWIG, H., HAINES, H. G., BISWAL, N. & BENYESH-MELNICK, M. (1972). The characterization of varicella-zoster virus DNA. *Journal of General Virology* 14, 111-114.
- MCGEOCH, D. J. & DAVISON, A. J. (1986a). Alpha herpesviruses possess a gene homologous to the protein kinase gene family of eukaryotes and retroviruses. *Nucleic Acids Research* 14, 1765-1777.
- MCGEOCH, D. J. & DAVISON, A. J. (1986b). DNA sequence of the herpes simplex virus type 1 gene encoding glycoprotein gH, and identification of homologues in the genome of varicella-zoster virus and Epstein-Barr virus. *Nucleic Acids Research* 14, 4281-4292.

- McGEOCH, D. J., DOLAN, A., DONALD, S. & RIXON, F. J. (1985). Sequence determination and genetic content of the short unique region in the genome of herpes simplex virus type 1. *Journal of Molecular Biology* **181**, 1-13.
- McGEOCH, D. J., DOLAN, A., DONALD, S. & BRAUER, D. H. K. (1986a). Complete DNA sequence of the short repeat region in the genome of herpes simplex virus type 1. *Nucleic Acids Research* **14**, 1727-1745.
- McGEOCH, D. J., DOLAN, A. & FRAME, M. C. (1986b). DNA sequence of the region in the genome of herpes simplex virus type 1 containing the exonuclease gene and neighbouring genes. *Nucleic Acids Research* **14**, 3435-3448.
- McKNIGHT, S. L. (1980). The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. *Nucleic Acids Research* **8**, 5949-5964.
- McLAUCHLAN, J. & CLEMENTS, J. B. (1983). DNA sequence homology between two co-linear loci on the HSV genome which have different transforming abilities. *EMBO Journal* **2**, 1953-1961.
- MARTIN, J. H., DOHNER, D. E., WELLINGHOFF, W. J. & GELB, L. D. (1982). Restriction endonuclease analysis of varicella-zoster vaccine virus and wild-type DNAs. *Journal of Medical Virology* **9**, 69-76.
- MESSING, J. (1979). A multi-purpose cloning system based on the single-stranded DNA bacteriophage M13. *Recombinant DNA Technical Bulletin* **2**, 43-48.
- MESSING, J. & VIEIRA, J. (1982). A new pair of M13 vectors for selecting either strand of double-digest restriction fragments. *Gene* **19**, 269-276.
- MISHRA, L., DOHNER, D. E., WELLINGHOFF, W. J. & GELB, L. D. (1984). Physical maps of varicella-zoster virus DNA derived with 11 restriction enzymes. *Journal of Virology* **50**, 615-618.
- MOUNT, S. M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Research* **10**, 459-472.
- MURCHIE, M.-J. & McGEOCH, D. J. (1982). DNA sequence analysis of an immediate-early gene region of the herpes simplex virus type 1 genome (map coordinates 0.950 to 0.978). *Journal of General Virology* **62**, 1-15.
- OAKES, J. E., ILLIS, J. P., HYMAN, R. W. & RAPP, F. (1977). Analysis by restriction enzyme cleavage of human varicella-zoster virus DNAs. *Virology* **82**, 353-361.
- OSTROVE, J. M., REINHOLD, W., FAN, C.-H., ZORN, S., HAY, J. & STRAUS, S. E. (1985). Transcription mapping of the varicella-zoster virus genome. *Journal of Virology* **56**, 600-606.
- PRESTON, V. G. & FISHER, F. B. (1984). Identification of the herpes simplex virus type 1 gene encoding the dUTPase. *Virology* **138**, 58-68.
- PUSTELL, J. & KAFATOS, F. C. (1982). A high speed, high capacity homology matrix: zooming through SV40 and polyoma. *Nucleic Acids Research* **10**, 4765-4782.
- QUINN, J. P. & McGEOCH, D. J. (1985). DNA sequence of the region in the genome of herpes simplex virus type 1 containing the genes for DNA polymerase and the major DNA binding protein. *Nucleic Acids Research* **13**, 8143-8163.
- RICHARDS, J. C., HYMAN, R. W. & RAPP, F. (1979). Analysis of the DNAs from seven varicella-zoster virus isolates. *Journal of Virology* **32**, 812-821.
- RIXON, F. J. & CLEMENTS, J. B. (1982). Detailed structural analysis of two spliced HSV-1 immediate-early mRNAs. *Nucleic Acids Research* **10**, 2241-2256.
- RIXON, F. J. & McGEOCH, D. J. (1984). A 3'-coterminal family of mRNAs from the herpes simplex virus type 1 short region: two overlapping reading frames encode unrelated polypeptides one of which has a highly reiterated amino acid sequence. *Nucleic Acids Research* **12**, 2473-2487.
- ROBERTS, T. M., SWANBERG, S. L., POTEETE, A., RIEDEL, G. & BACKMAN, K. (1980). A plasmid cloning vehicle allowing a positive selection for inserted fragments. *Gene* **12**, 123-127.
- SANGER, F., COULSON, A. R., BARRELL, B. G., SMITH, A. J. H. & ROE, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *Journal of Molecular Biology* **143**, 161-178.
- STADEN, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Research* **10**, 4731-4751.
- STADEN, R. & McLACHLAN, A. D. (1982). Codon preference and its use in identifying protein coding regions in long DNA sequences. *Nucleic Acids Research* **10**, 141-156.
- STOW, N. D. & DAVISON, A. J. (1986). Identification of a varicella-zoster virus origin of DNA replication and its activation by herpes simplex virus type 1 gene products. *Journal of General Virology* **67**, 1613-1623.
- STRAUS, S. E., AULAKH, H. S., RUYECHAN, W. T., HAY, J., CASEY, T. A., VANDE WOUDE, G. F., OWENS, J. & SMITH, H. A. (1981). Structure of varicella-zoster virus DNA. *Journal of Virology* **40**, 516-525.
- STRAUS, S. E., OWENS, J., RUYECHAN, W. T., TAKIFF, H. E., CASEY, T. A., VANDE WOUDE, G. F. & HAY, J. (1982). Molecular cloning and physical mapping of varicella-zoster virus DNA. *Proceedings of the National Academy of Sciences, U.S.A.* **79**, 993-997.
- STRAUS, S. E., HAY, J., SMITH, H. & OWENS, J. (1983). Genome differences among varicella-zoster virus isolates. *Journal of General Virology* **64**, 1031-1041.
- TAMURA, T., NODA, M. & TAKANO, T. (1981). Structure of the baboon endogenous virus genome: nucleotide sequences of the long terminal repeat. *Nucleic Acids Research* **9**, 6615-6626.
- TAYLOR, P. (1984). A fast homology program for aligning biological sequences. *Nucleic Acids Research* **12**, 447-455.
- TAYLOR, P. (1986). A computer program for translating DNA sequences into protein. *Nucleic Acids Research* **14**, 437-441.
- TOMITA, M. & MARCHESI, V. T. (1975). Amino acid sequence and oligosaccharide attachment sites of human erythrocyte glycoprotein. *Proceedings of the National Academy of Sciences, U.S.A.* **72**, 2964-2968.
- VIEIRA, J. & MESSING, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 259-268.
- WAGNER, E. K. (1985). Individual HSV transcripts. Characterization of specific genes. In *The Herpesviruses*, vol. 3, pp. 45-104. Edited by B. Roizman. New York: Plenum Press.

- WATSON, R. J. & VANDE WOUDE, G. F. (1982). DNA sequence of an immediate-early gene (IE mRNA-5) of herpes simplex virus type 1. *Nucleic Acids Research* **10**, 979-991.
- WATSON, R. J., SULLIVAN, M. & VANDE WOUDE, G. F. (1981*a*). Structures of two spliced herpes simplex virus type 1 immediate-early mRNAs which map at the junctions of the unique and reiterated regions of the virus DNA S component. *Journal of Virology* **37**, 431-444.
- WATSON, R. J., UMENE, K. & ENQUIST, L. W. (1981*b*). Reiterated sequences within the intron of an immediate-early gene of herpes simplex virus type 1. *Nucleic Acids Research* **9**, 4189-4199.
- WELLER, S. K., SPADARO, A., SCHAFFER, J. E., MURRAY, A. W., MAXAM, A. M. & SCHAFFER, P. A. (1985). Cloning, sequencing, and functional analysis of ori_L, a herpes simplex virus type 1 origin of DNA synthesis. *Molecular and Cellular Biology* **5**, 930-942.
- ZWEERINK, H. J., MORTON, D. H., STANTON, L. W. & NEFF, B. J. (1981). Restriction endonuclease analysis of the DNA from varicella-zoster virus: stability of the DNA after passage *in vitro*. *Journal of General Virology* **55**, 207-211.

(Received 7 April 1986)